

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2003, 19:37:10 ; Search time 69 Seconds
(without alignments)
549.459 Million cell updates/sec

Title: US-09-867-753-2
Perfect score: 986
Sequence: 1 MARSLVHDTVFYCLSVYQVK.....LMLANELRADPDDCVYIVVD 184

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

```
Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%
	Listing first	45 summaries

```

1: SP:archaea:*
2: SP:bacteria:*
3: SP:fungi:*
4: SP:human:*
5: SP:invertebrate:*
6: SP:mammal:*
7: SP:mhc:*
8: SP:organelle:*
9: SP:phage:*
10: SP:plant:*
11: SP:rodent:*
12: SP:virus:*
13: SP:unclassified:*
14: SP:unclassified:*
15: SP:ivrus:*
16: SP:bacteriap:*
17: SP:archaeap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	789	80.0	148	4	Q95030	095030 homo sapien
2	260.5	26.4	137	4	Q954M1	Q954M1 homo sapien
3	216	21.9	288	4	Q9BR00	Q9BR00 homo sapien
4	216	21.9	288	4	Q9BQV4	Q9BQV4 homo sapien
5	212.5	21.6	227	11	Q9EQM5	Q9EQM5 mus musculus
6	212	21.5	286	5	Q87615	Q87615 branchiosto
7	197	20.0	382	11	Q88933	Q88933 mus musculus
8	197	20.0	387	11	Q922U3	Q922U3 mus musculus
9	196	19.9	314	11	Q54817	Q54817 mus musculus
10	191.5	19.4	248	13	Q12952	Q12952 cynops pyrri
11	191.5	19.4	562	4	Q960S3	Q960S3 homo sapien
12	190	19.3	640	5	Q8TDM4	Q8TDM4 drosophila
13	188.5	19.1	371	5	Q61282	Q61282 hemiceutrot
14	188	19.1	327	11	Q70238	Q70238 mus musculus
15	187	19.0	328	13	Q9W7M5	Q9W7M5 brachydania
16	186	18.9	563	5	Q9VTK6	Q9VTK6 drosophila

17	183	8.6	408	5	09VPI1	09VPI1 drosophila
18	182.5	18.5	185	13	P79857	P79857 pleurodeles
19	181	18.4	365	6	09GMA3	09GMA3 bos taurus
20	180	18.3	240	11	08R4I3	08R4I3 mus musculus
21	180	18.3	229	6	08S003	08S003 canis famli
22	180	18.3	299	11	09WYQ9	09WYQ9 rattus norv
23	180	18.3	299	11	09JLY8	09JLY8 rattus norv
24	179.5	18.2	362	5	021836	021836 caenorhabd1
25	179	18.2	301	5	046170	046170 tribolium c
26	179	18.2	391	5	08T8C1	08T8C1 gryllus bim
27	179	18.2	826	5	09B130	09B130 clona intes
28	177.5	18.0	276	13	098T97	098T97 gallus galli
29	177	18.0	387	13	093582	093582 gallus galli
30	176.5	17.9	493	11	09SE16	09SE16 mus musculus
31	176	17.8	214	5	08UYD3	08UYD3 gallus galli
32	176	17.8	370	5	025411	025411 lineus sang
33	176	17.8	479	11	09CXI6	09CXI6 mus musculu
34	176	17.8	479	13	013081	013081 coturnix co
35	176	17.8	612	5	024477	024477 drosophila
36	176	17.8	612	5	026441	026441 drosophila
37	176	17.8	835	4	096H85	096H85 homo sapien
38	175.5	17.8	328	5	062546	062546 hydra aten
39	175.5	17.8	371	5	046169	046169 tribolium c
40	175	17.7	484	11	09CCK7	09CCK7 mus musculu
41	174.5	17.7	350	13	09JAL2	09JAL2 gallus galli
42	174	17.6	282	13	073678	073678 oryzaas lab
43	174	17.6	465	5	096824	096824 hemitecentrot
44	174	17.6	294	5	09NDA9	09NDA9 branchiosto
45	173	17.5	210	11	09QYR0	09QYR0 stochomys l

ALIGNMENTS

	RESULT	1
095030	ID	095030
	PRELIMINARY:	PRT: 148 AA.
AC	095030:	
DT	01-MAY-1999	(TREMBLrel. 10, Created)
DT	01-MAY-1999	(TREMBLrel. 10, Last sequence update)
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)
DE	WUOSSC:H.GS421103.1 protein (Fragment).	
GN	WUOSSC:H.GS421103.1	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
CC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Leonard S., Graves T., Coffman M.;	
RT	"The sequence of Homo sapiens BAC clone GS1-42113.";	
RL	Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.	
RN	[2]	
RN	SEQUENCE FROM N.A.	
RP	Waterston R.;	
RL	Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.	
DR	EMBL; AC005023; AAC78617.1; -	
DR	HSSP; P06601; IFJL.	
DR	InterPro; IPR001356; Homeobox.	
DR	Pfam; PF00046; homeobox; 1.	
DR	SMART; SM00389; HOX; 1.	
DR	PROSITE; PS00071; HOMEBOX_2; 1.	
SO	NON TER	148
SEQUENCE	148 AA; 16143 MW; 7C81BD318E70825C CXC64;	
Query Match	80.0%; Score 789; DB 4; Length 148;	
Best Local Similarity	100.0%; Pred. No. 6.3e-63;	
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY	1 MARSLVHDTVFYCLSVYQKISPTPOLGAASSAEHHVGAGAGLCGNNPEGCVNHENGM 60	
DB	1 MARSLVHDTVFYCLSVYQKISPTPOLGAASSAEHHVGAGAGLCGNNPEGCVNHENGM 60	

```
OY 61 NRDSGMIEGGGNOEPPOPPPEEPPOAAMSGPOENMOPTRTKFTLLQVEEELS 120
DB 61 NRDSGMIEGGGNOEPPOPPPEEPPOAAMSGPOENMOPTRTKFTLLQVEEELS 120
OY 121 VFRHTQYDPVPTFRRLAENLGVTEDEKVR 148
DB 121 VFRHTQYDPVPTFRRLAENLGVTEDEKVR 148

RESULT 2
OY4W1 PRELIMINARY: PRT: 137 AA.
AC O9Y4W1;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE D5J13M9.1 (Novel homeobox domain protein) (Fragment).
GN D5J13M9.1
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bird C.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AL049631; CAB46720.1; -.
DR HSSP; P06601; 1FJL.
DR InterPro; IPR001356; HTH_repressr.
DR InterPro; IPR000047; HTH_repressr.
DR InterPro; IPR000327; POU_domain.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR PRINTS; PR00031; HTHREPRESSR.
DR PRODOM; PD000028; POU_DOMAIN.
DR PRODOM; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
FT NON_TER 1
FT TER 137
SQ SEQUENCE 137 AA; 16126 MW; 9B2B9CF358E59A8C CRC64;

Query Match 26.4%; Score 260.5; DB 4; Length 137;
Best Local Similarity 47.5%; Pred. No. 7.9e-16;
Matches 56; Conservative 15; Mismatches 28; Indels 19; Gaps 3;

OY 76 EPROQOPPP-----EPPAQAAMEGPP-----ENNQP-----RTTRTKFTLLQVE 116
DB 15 KPEQOQEEPLLEKQOEDEPPQTVESGPQPAEGPQTAEGPQPRKRKRRTTAFTQFOLO 74

OY 117 ELESVFRHTQYDPVPTFRRLAENLGVTEDEKVRVFNKRRARCRHORELMLANELRAD 174
DB 75 ELENFEDESQPDVYARRLAARLNLTEDRVQVWQNRRAKMKRNRQVLMRLNTATAD 132

RESULT 3
O9BR00 PRELIMINARY: PRT: 288 AA.
AC O9BR00;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DE 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Hypothetical 31.6 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Howell G.R., Huckle E., Ross M.T.;

OY 126 QYDPVPTFRRLAENLGVTEDEKVRVFNKRRARCRHORELMLANEL 171
DB 99 GSDGNVEDSDSEKEPPGOYSRPGAVGLEPNAQOPVH--ATPPLQLELECFQRE 156
OY 157 QPSEFLRRLARLSMNVTELAVQITFENRRKMRKHQALMARNL 202

RESULT 4
O9BOY4 PRELIMINARY: PRT: 288 AA.
AC O9BOY4;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DE 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Hypothetical 31.7 kDa protein (CDNA FLJ25356 fis, clone TST02591)
DE (Homeobox protein) (Homeobox protein from AL590526).
GN THG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Howell G.R., Huckle E., Ross M.T.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Hoshita T., Hiraoaka S., Murakawa K., Takiguchi S., Kusano J.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Suzuki Y., Hata H., Nakagawa K., Mizuno S., Koriinaga M., Kawamura M.,
RA Kawakami B., Nagai K., Isogai T., Sato H., Nishikawa T., Sugiyama A.,
RA "NEDO human cDNA sequencing project."
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Yin Y., Jin Y., Levine A.J.;
RT "Molecular cloning and characterization of a novel testis homeobox
RL gene, THG1, as a potential testicular tumor suppressor gene."
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
```

```
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AL590524; CAC36517.1; -.
DR HSSP; P06601; 1FJL.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR001993; Mitoch_carrier.
DR Pfam; PF00046; homeobox; 1.
DR PRODOM; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
DR PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
KW DNA-binding; Homeobox; Hypothetical protein; Nuclear protein.
SQ SEQUENCE 288 AA; 31637 MW; 50B571B13DB712B9 CRC64;

Query Match 21.9%; Score 216; DB 4; Length 288;
Best Local Similarity 34.9%; Pred. No. 1.7e-11;
Matches 58; Conservative 23; Mismatches 57; Indels 28; Gaps 7;

OY 23 PTPQGAASSAEGHVGAGPGLMGNMNEGGVNHENGNRDG--MP-----E 69
DB 48 PEPQGTAAEGK---LKSAGAGGEEKDGC---GEEKDGGAGVGHLMEGNLEGSTS 98

OY 70 GGGGNOEPPOPPPEEP---AQAMSGPOEN--MOPTRTKFTLLQVEEESVFRHT 125
DB 99 GSDGNVEDSDSEKEPPGOYSRPGAVGLEPNAQOPVH--ATPPLQLELECFQRE 156

OY 126 QYDPVPTFRRLAENLGVTEDEKVRVFNKRRARCRHORELMLANEL 171
DB 99 GSDGNVEDSDSEKEPPGOYSRPGAVGLEPNAQOPVH--ATPPLQLELECFQRE 156
OY 157 QPSEFLRRLARLSMNVTELAVQITFENRRKMRKHQALMARNL 202

RESULT 4
O9BOY4 PRELIMINARY: PRT: 288 AA.
AC O9BOY4;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DE 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Hypothetical 31.7 kDa protein (CDNA FLJ25356 fis, clone TST02591)
DE (Homeobox protein) (Homeobox protein from AL590526).
GN THG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Howell G.R., Huckle E., Ross M.T.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Hoshita T., Hiraoaka S., Murakawa K., Takiguchi S., Kusano J.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Suzuki Y., Hata H., Nakagawa K., Mizuno S., Koriinaga M., Kawamura M.,
RA Kawakami B., Nagai K., Isogai T., Sato H., Nishikawa T., Sugiyama A.,
RA "NEDO human cDNA sequencing project."
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Yin Y., Jin Y., Levine A.J.;
RT "Molecular cloning and characterization of a novel testis homeobox
RL gene, THG1, as a potential testicular tumor suppressor gene."
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
```

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL: AL590526; CAC36519.1; -;
 DR EMBL: AK058125; BAB71675.1; -;
 DR EMBL: AF317219; AAL02160.1; -;
 DR EMBL: BC021719; AAH21719.1; -;
 DR HSSP: P06601; 1FJL.
 DR InterPro: IPR001356; Homeobox.
 DR InterPro: IPR001933; Mitoch_carrier.
 DR Pfam: PF00046; homeobox; 1.
 DR ProDom: PD000010; Homeobox; 1.
 DR SMART: SM00389; HOX; 1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS00071; HOMEBOX_2; 1.
 DR PROSITE: PS00215; MITOCH_CARRIER; UNKNOWN; 1.
 DR DNA-binding: Homeobox; Hypothetical protein; Nuclear protein.
 KW SEQUENCE 288 AA; 31691 MW; 280E007064515808 CRC64;
 SQ

Query Match 21.9%; Score 216; DB 4; Length 288;
 Best Local Similarity 34.9%; Pred. No. 1.7e-11;
 Matches 58; Conservative 23; Mismatches 57; Indels 28; Gaps 7;

OY 23 PPPLGASASAGHVGQAPGLMGNMPEGVYHENGNNRDG--MIP-----E 69
 DB 48 PPEPGTAGER---LKSAGAGGEGEKDG-----GEEKDGAGVPGHLMGDEGTS 98
 OY 70 GGGGNEPRQOPPEEP---AQAMEGPOPEM--MOPTRKRTKFTLLQVEELSFRRHT 125
 DB 99 GSDGVEDSDQSEKPEGGQYSRPGAVGLPEPNAOQPVH--APPLDLOELERIFORE 156
 OY 126 QYPDVPTRELAENLGVTEDEVKRVWFKNRARCRRHORELMLANEL 171
 DB 157 QPPEELRRRLARSMVTELAVQIWFENRRAKRRRQRLAMRNL 202

RESULT 5 Q9EQM5 PRELIMINARY: PRT; 227 AA.

AC Q9EQM5;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Homeobox protein GPBOX.
 GN 160002601R1K OR GPBOX.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Takasaki N., McIsaac R., Dean J.;
 RT "Gbox, a novel homeobox gene preferentially expressed in female germ
 RT cells at the onset of sexual dimorphism in mice."
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF201698; AAG36768.1; -;
 DR MGD: MGI:1925663; 1600026001R1K.
 DR InterPro: IPR001356; Homeobox.
 DR Pfam: PF00046; homeobox; 1.
 DR ProDom: PD000010; Homeobox; 1.
 DR SMART: SM00389; HOX; 1.
 DR PROSITE: PS00071; HOMEBOX_2; 1.
 DR PROSITE: PS00071; HOMEBOX_2; 1.
 SQ SEQUENCE 227 AA; 25112 MW; D39CF6795AF682DC CRC64;

Query Match 21.6%; Score 212.5; DB 11; Length 227;
 Best Local Similarity 35.7%; Pred. No. 2.7e-11;
 Matches 61; Conservative 17; Mismatches 60; Indels 33; Gaps 6;

OY 28 GAASSAEHVGQAPGLMGNMNP-----EGVYHENGNNRDG----- 66
 DB 54 GGLDGEAGGAVAGGEAQEPAPLSPQENTGGESEENEGEGHAGDAGSPED 113
 OY 67 --IPEGGNG--OEPPOQPPPEPAQAMGPOPEM--MOPTRRTKFTLLQVEELS 120
 DB 114 DNIEGEGGGINIQPPQQAALPE-----GMRNPAGNLTARQTRRTFTISQLADLER 168

OY 121 VERNHPOVPTRELAENLGVTEDEVKRVWFKNRARCRRHORELMLANEL 171
 DB 169 LFGENRPPSLVRRLARLAWMGVDSVQEWFKRRALFRHSH-LMFCFL 218

RESULT 6

OY 08T615 PRELIMINARY: PRT; 286 AA.
 AC 08T615;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Paired-like homeodomain transcription factor Shox.
 GN SHOX.
 OS Branchiostoma floridae (Florida lancelet) (Amphioxus).
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 OC Branchiostoma.
 OX NCBI_TaxID=7739;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Jackman W.R., Jr., Kimmel C.B.;
 RT "Coincident iterated gene expression in the amphioxus neural tube."
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF465939; AAL83210.1; -;
 KW Homeobox; DNA-binding; Nuclear protein.
 SQ SEQUENCE 286 AA; 31551 MW; 6BF50B62D92958C1 CRC64;

Query Match 21.5%; Score 212; DB 5; Length 286;
 Best Local Similarity 32.9%; Pred. No. 3.9e-11;
 Matches 52; Conservative 24; Mismatches 48; Indels 34; Gaps 5;

OY 28 GAASSAEH--VGQAPGLMGNMPEGVYHENGNNRDGMIPEG--GGGNEP---RQOP 81
 DB 46 GAEVDVGDGLCPVSGPGL-----FVGGGGENSGPENKDP 82
 OY 82 QPPEEPAQAAMGPOPEMNPQPTRRKFTLLQVEELSFRRHPOVPTRELAENIG 141
 DB 83 SKTGEKKDLSPNSKDGKIKORRSTRNFTLEQLERLFDTHYPDAFMRELSORUG 142
 OY 142 VTEDEVKRVWFKNRARCRRHOREL-----MLANELRAD 174
 DB 143 LSEARVQVWFQNNRACRCRQENOLQKGMICSLALGN 180

RESULT 7 O88933 PRELIMINARY: PRT; 382 AA.

AC O88933;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Homeobox protein SPXL.
 GN ESXL OR EPX.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=97398449; PubMed=9256347;
 RA Branford W.W., Zhao G.O., Valerius M.T., Weinstein M.,
 RA Birkenmeier E.H., Rowe L.B., Potter S.S.;
 RT "Spxl, a novel X-linked homeobox gene expressed during
 RT spermatogenesis."
 RL Mech. Dev. 65:87-98(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=97392770; PubMed=9245514;
 RA Li Y., Lemaire P., Behringer R.R.;
 RT "Esxl, a novel X chromosome-linked homeobox gene expressed in mouse
 RT extraembryonic tissues and male germ cells."
 RL Dev. Biol. 188:85-95(1997).

RN [3]
 RP SEQUENCE FROM N.A.
 RA Branford M.W., Potter S.S.;
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV;
 RA Yan Y.-T., Yang L., Sciavolino P.J., Wang H., Chan D.C.,
 RA Abate-Shen C., Shen M.M.;
 RT "Pxx: a novel paired-like homeobox gene expressed in the chorion and
 placenta.";
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL: AF085715; AAC5366.1; -;
 DR EMBL: AF017735; AAD01622.1; -;
 DR HSSP: P06601; IFTL.
 DR TRANSFAC: T03478; -;
 DR MGD: MGI:1096388; Esxl.
 DR InterPro: IPR001356; Homeobox.
 DR InterPro: IPR000047; HTH_repressr.
 DR Pfam: PF00046; homeobox.1.
 DR PRINTS: PR00031; HTHREPRESSR.
 DR PRODOM: PD000010; Homeobox.1.
 DR SMART: SM00389; HOX.1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS50071; HOMEBOX_2; 1.
 DR DNA-binding: Homeobox; Nuclear protein.
 SQ SEQUENCE 382 AA; 43540 MW; EC92301A84DA6175 CRC64;

Query Match 20.0%; Score 197; DB 11; Length 382;
 Best Local Similarity 34.1%; Pred. No. 1.2e-09;
 Matches 59; Conservative 18; Mismatches 72; Indels 24; Gaps 5;
 QY 17 YQVKSIPTPQGAASSAGHYGCGAPGLMGNNPBGVNHNGMNRDGG----- 65
 AC 054817; PRELIMINARY; PRT; 387 AA.
 DB 84 YQEBEGFEPSRGEAAAP--VAEAPQAWNGNENIGGFLSENAOLGENDAAPVROSILMP 140
 QY 66 -MIEPGGGGNO-----EPPOQPP-----PEEPAQAMEGPPQENMQPTRTKFTLLQVE 116
 DB 141 LMQVPAOSSPQPLRANPLQAPQPEDEEEDQEGEPOQOE-PKPRRYICFTPILOQ 199
 QY 117 ELESVFRHTQYPDVPTRELAENIGVTEDKYRVWFKNRKRCRRHORELMLAN 169
 DB 200 ELEAFQFQVQYPDLEFARVELARILGLPEPRVQVWFQNRRAKWRRLRAQAFRN 252

RESULT 8
 ID 092203 PRELIMINARY; PRT; 387 AA.
 AC 092203;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Homeodomain protein Epx (Fragment).
 GN Esxl OR epx.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV;
 RX MEDLINE=20079279; PubMed=10611245;
 RA Yan Y.T., Stein S.M., Ding J., Shen M.M., Abate-Shen C.;
 RT "A Novel PF/PN Motif Inhibits Nuclear Localization and DNA Binding
 Activity of the Esxl Homeoprotein.";
 RL Mol. Cell. Biol. 20:661-671(2000).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL: AF017734; AAD01621.1; -;
 DR HSSP: P06601; IFTL.
 DR MGD: MGI:1096388; Esxl.
 DR InterPro: IPR001356; Homeobox.
 DR InterPro: IPR000047; HTH_repressr.
 DR InterPro: IPR000047; HTH_repressr.

DR InterPro: IPR002965; P-rich_extensn.
 DR Pfam: PF00046; homeobox.1.
 DR PRINTS: PR00031; HTHREPRESSR.
 DR PRINTS: PR01217; PRICHEXTENSN.
 DR PRODOM: PD000010; Homeobox.1.
 DR SMART: SM00389; HOX.1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS50071; HOMEBOX_2; 1.
 DR DNA-binding: Homeobox; Nuclear protein.
 FT NOW TRER
 SQ SEQUENCE 387 AA; 44170 MW; 0D6A6E4C36CF9CD3 CRC64;

Query Match 20.0%; Score 197; DB 11; Length 387;
 Best Local Similarity 34.1%; Pred. No. 1.2e-09;
 Matches 59; Conservative 18; Mismatches 72; Indels 24; Gaps 5;
 QY 17 YQVKSIPTPQGAASSAGHYGCGAPGLMGNNPBGVNHNGMNRDGG----- 65
 DB 89 YQEBEGFEPSRGEAAAP--VAEAPQAWNGNENIGGFLSENAOLGENDAAPVROSILMP 145
 QY 66 -MIEPGGGGNO-----EPPOQPP-----PEEPAQAMEGPPQENMQPTRTKFTLLQVE 116
 DB 146 LMQVPAOSSPQPLRANPLQAPQPEDEEEDQEGEPOQOE-PKPRRYICFTPILOQ 204
 QY 117 ELESVFRHTQYPDVPTRELAENIGVTEDKYRVWFKNRKRCRRHORELMLAN 169
 DB 205 ELEAFQFQVQYPDLEFARVELARILGLPEPRVQVWFQNRRAKWRRLRAQAFRN 257

RESULT 9
 ID 054817 PRELIMINARY; PRT; 314 AA.
 AC 054817;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Paired-like homeodomain containing protein.
 GN Esxl OR Etx.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SWISS;
 RX MEDLINE=97392770; PubMed=9245514;
 RA Li Y., Lemaire P., Behringer R.R.;
 RT "Esxl, a novel X chromosome-linked homeobox gene expressed in mouse
 extraembryonic tissues and male germ cells.";
 RL Dev. Biol. 188:85-95(1997).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL: AF004211; AAB94670.1; -;
 DR HSSP: P06601; IFTL.
 DR TRANSFAC: T03474; -;
 DR MGD: MGI:1096388; Esxl.
 DR InterPro: IPR001356; Homeobox.
 DR InterPro: IPR000047; HTH_repressr.
 DR Pfam: PF00046; homeobox.1.
 DR PRINTS: PR00031; HTHREPRESSR.
 DR PRODOM: PD000010; Homeobox.1.
 DR SMART: SM00389; HOX.1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS50071; HOMEBOX_2; 1.
 DR DNA-binding: Homeobox; Nuclear protein.
 SQ SEQUENCE 314 AA; 35969 MW; 7256F81D4AA246 CRC64;

Query Match 19.9%; Score 196; DB 11; Length 314;
 Best Local Similarity 34.1%; Pred. No. 1.2e-09;
 Matches 59; Conservative 18; Mismatches 72; Indels 24; Gaps 5;
 QY 17 YQVKSIPTPQGAASSAGHYGCGAPGLMGNNPBGVNHNGMNRDGG----- 65
 DB 16 YQEBEGFEPSRGEAAAP--VAEAPQAWNGNENIGGFLSENAOLGENDAAPVROSILMP 72

Best Local Similarity 32.28; Pred. No. 9.1e-09;
Matches 55; Conservative 20; Mismatches 60; Indels 36; Gaps 7;

QY 24 TPOLGAASSAGHGOGAPGLM-----GNM--NP---EGGVHNGMNDGC----- 65
AC 061282; PRELIMINARY; PRT; 371 AA.
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Orthodenticle-related protein.
GN HPOPTX.
OS Hemiteutrobus pulcherrimus (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinozoa; Euechinozoa; Echinacea; Echinozoa; Strongylocentrotidae;
OC Hemiteutrobus.
OC NCB1_Taxid=7650;
RN [1]
RP SEQUENCE FROM N.A.
RA Shimada H., Akasaka K., Mitsuana-Nakatsubo K., Takata K.,
RA Sakamoto N.,
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97166044; PubMed=9013937;
RA Sakamoto N., Akasaka K., Mitsuana-Nakatsubo K., Takata K.,
RA Nishitani T., Shimada H.,
RT "Two isoforms of orthodenticle-related proteins (HPOtx) bind to the
RT enhancer element of sea urchin arylsulphatase gene."
RL Dev. Biol. 181:284-295(1997).
CC -I- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL: AB011526; BAA28675.1; -.
DR HSPSP; P06601; IFTL.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox; 1.
DR PRINTS: PR00024; HOMEBOX.
DR PRODOM: PD000010; Homeobox; 1.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS00071; HOMEBOX_2; 1.
DR DNA-binding: Homeobox; Nuclear protein.
SQ SEQUENCE 371 AA; 41272 MW; 6F4BEF71251E5EF7 CRC64;

Query Match 19.18; Score 188.5; DB 5; Length 371;
Best Local Similarity 28.98; Pred. No. 6.6e-09;
Matches 52; Conservative 34; Mismatches 53; Indels 41; Gaps 8;

QY 14 LSYGVKRSPTPOLGAASSAE-----GHVGOGAPGLMGNPNPESGVN----- 55
DB 25 LNVAPMKLE---RVGMSSSPRLITDCNTRSP--VPSHMEPPGAGRVPMILYPRQY 79
QY 56 -HENGMMNDGCM-----IPEGGGGNOEPPOPO-----PPEEPAQAAMEGPOENMPOR 104
DB 80 AYSNPMYGEGLAPADRVPP---TQGHMFGQPVYLGMTSEKSHNSKNDVDPKQO----R 132
QY 105 TRRKFTLLQVEELSEVFRHTQYDVPVTRRELAEVLGTEDEKRVWFKKRRACRRHQRE 164
DB 133 RERTTFRAQIDVLETLEFSRTYRPIDEFMRREVAMKINLPESRVQVWFKRRAKRACROOQOO 192

RESULT 14
ID 070238 PRELIMINARY; PRT; 227 AA.

AC 070238;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Homeobox protein PSX.
GN PSX.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCB1_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98172748; PubMed=9511757;
RA Han Y.J., Park A.R., Sung D.Y., Chun J.Y.;
RT "Pax, a novel murine homeobox gene expressed in placenta."
RL Gene 207:159-166(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Han Y.J., Park A.R., Seong D.Y., Chun J.Y.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF017453; AAD08781.1; -.
DR HSPSP; P06601; IFTL.
DR TRANSFAC; T03294; -.
DR MGD; MGI:1202888; Psx1.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox; 1.
DR PRODOM: PD000010; Homeobox; 1.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00071; HOMEBOX_2; 1.
SQ SEQUENCE 227 AA; 25484 MW; 0FEF5EC1C5EAE374 CRC64;

Query Match 19.18; Score 188; DB 11; Length 227;
Best Local Similarity 33.18; Pred. No. 4.2e-09;
Matches 55; Conservative 22; Mismatches 57; Indels 32; Gaps 5;

QY 28 GAASSAESHGOGAPGLMGNPNP-----EGGVHNGMNDGCM----- 66
DB 54 GEFDDGGEACQAEVAGGEQAOEPAFLSPAQATGGEDEGENKEGEMGRHAGDGAASSED 113
QY 67 --IPEGGGGN--OEPRPOPPPEEPAQAAMEGPOEN--MOPRTTRFTLLQVEELSE 120
DB 114 DSILEEGGENIDQPPQQAASPD-----STRNPVNLRLAQLKRRRTFTHSQHLDER 168
QY 121 VFRHTQYDVPVTRRELAEVLGTEDEKRVWFKKRRACRRHQRELM 166
DB 169 LFOETRYPSLARRLARWAGVDECDVGNWFRMRALQRRNRVILM 214

RESULT 15
ID 09W7M5 PRELIMINARY; PRT; 328 AA.
AC 09W7M5;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Orthopedia protein.
GN ORP OR ORP.
OS Brachydonto rerio (zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OC NCB1_Taxid=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Del Giacco L., Di Benedetto B., Duga S., Dianl S., Cotelletti F.,
RT "Isolation of the mRNA encoding Otp (Orthopedia) in the zebrafish,
RT Danio rerio."
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -I- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL: AF071496; AAD42021.1; -.

This Page Blank (uspto)

GenCore version 5.1.4.D5_A578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2003, 19:37:10 ; Search time 37 seconds

(without alignments)
478.074 Million cell updates/sec

Title: US-09-867-753-2

Perfect score: 986
Sequence: 1 MARSLVHTVYCYCLSVYQV.....LMLANELRADPDCCYIYVD 184

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	189.5	19.2	399	2 JC6522	paired-type homeod
2	185.5	18.8	252	1 A54677	homeotic protein g
3	185.5	18.8	256	1 A42768	homeotic protein g
4	184.5	18.7	217	2 I48902	homeobox protein P
5	184.5	18.7	245	2 S26076	homeotic protein K
6	183	18.6	384	2 A46403	transcription fact
7	181	18.4	419	1 T70617	homeotic protein g
8	179.5	18.2	362	2 T24046	hypothetical prote
9	178	18.1	185	2 A55882	homeobox protein H
10	177	18.0	326	2 A45452	transcription fact
11	176.5	17.9	290	2 I49265	paired box transcr
12	176	17.8	245	2 I51226	homeodomain protei
13	176	17.8	247	1 UC6540	placenta specific-
14	176	17.8	280	2 I48713	Phox2 homeodomain
15	176	17.8	370	2 JC6130	paired box transcr
16	176	17.8	479	1 S15031	paired box transcr
17	176	17.8	612	2 A54282	reversed polarity
18	174.5	17.7	245	1 A47539	homeotic protein g
19	173.5	17.6	333	1 JM0097	blood-related hom
20	173	17.5	80	2 A26332	homeotic protein B
21	173	17.5	346	1 I48185	gene alx3 protein
22	172	17.4	449	1 B43698	paired box transcr
23	172	17.4	520	2 S78502	paired box transcr
24	171.5	17.4	243	1 BA2768	homeotic protein g
25	171.5	17.4	243	2 I51424	homeotic protein g
26	171.5	17.4	427	1 A3698	paired box transcr
27	170.5	17.3	423	1 A26062	paired box segment
28	168.5	17.1	242	2 A43904	homeotic protein G
29	168.5	17.1	278	2 A56570	homeobox protein D

30	167.5	17.0	283	2 I50112	DLX4 homeodomain P
31	167	16.9	284	2 JS0659	homeotic protein H
32	166.5	16.9	284	2 S60250	mab-18 protein (tr
33	166	16.8	296	2 S60251	paired type homeob
34	166	16.8	314	2 JC5273	homeodomain protei
35	166	16.8	318	2 S52424	hypothetical protei
36	166	16.8	319	2 T18786	homeodomain protei
37	166	16.8	319	2 T43635	homeotic protein S
38	165.5	16.8	288	2 S27842	CVC domain-contain
39	165.5	16.8	363	2 JC7750	homeotic protein o
40	165.5	16.8	671	2 A35912	cartilage homeopro
41	165	16.7	326	2 A47523	homeotic protein P
42	165	16.7	798	2 S20881	paired box transcr
43	164	16.6	491	1 S60252	homeotic protein o
44	163.5	16.6	354	2 S39406	otx1 protein - mou
45	163.5	16.6	355	2 S35345	

ALIGNMENTS

```

RESULT 1
JC6522
paired-type homeodomain protein, Alx-4 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 31-Mar-2000
C:Accession: JC6522
R:Ou, S.; Li, L.; Wisdom, R.
Gene 203, 217-223, 1997
A:Title: Alx-4: cDNA cloning and characterization of a novel paired-type homeodomain
A:Reference number: JC6522; MUID:98086222; PMID:9426253
A:Contents: Embryo
A:Accession: JC6522
A:Molecule type: mRNA
A:Residues: 1-399 <OUA>
A:Cross-references: GB:AF001465; NID:q2352265; PIDN:AAC39943.1; PID:q2352266
C:Comment: This protein belongs to the family of paired-type homeodomain proteins, it
C:Genetics:
A:Gene: Alx-4
C:Superfamily: homeotic protein Hox B3; homeobox homology
C:Keywords: DNA binding; embryo; homeobox; nucleus; transcription regulation
F:203-259/Domain: homeobox homology <HOX>

Query Match      19.2%; Score 189.5; DB 2; Length 399;
Best Local Similarity 30.1%; Pred. No. 2.2e-07;
Matches 55; Conservative 22; Mismatches 49; Indels 57; Gaps 7;

OY 23 PPQ---LGAASAEHVGAGAPGLGMNMBEGVNHENGMNMDGGMPEGGGQNEPRO 79
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 95 PTPQPPAPAPPAHLYLORGA-----CKTPPDGSLK-----LDGSGGHNAALQ 139
OY 80 QP-----QPPPEPA-----QAAMEGRO-----PENMQ----- 102
Db 140 VPCYAKESNLGEPELPDSEPPYGMQMSYLVKETGAKGQDRASATIPLEKTEDESNK 199
OY 103 --PRTTRFTLLQVELESVRHNTQYPPVPTTRRELAEMLGVTEDKVRVFKKARCR 160
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 200 GKRRRRRTFTSYQLELEKVFQKTHYPDVYAREQDLAMRTDLTEARVQWFOHRAKMR 259
OY 161 HOR 163
    | |
Db 260 REF 262

RESULT 2
A54677
homeotic protein gooseoid [validated] - human
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Sep-2000
C:Accession: A54677
R:Blum, M.; De Robertis, E.M.; Kojis, T.; Heinzmann, C.; Klisak, I.; Geisler, D.; Sp
Genomics 21, 388-393, 1994
A:Title: Molecular cloning of the human homeobox gene gooseoid (GSC) and mapping of

```

```

A:Reference number: A54677; MUID:94375063; PMID:7916327
A:Accession: A54677
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-252 <BLU>
C:Genetics:
A:Gene: GDB:GSC
A:Cross-references: GDB:251683; OMIM:138890
C:Map position: 14q32.1-14q32.1
C:Superfamily: homeotic protein goosecoid; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:156-212/Domain: homeobox homology <HOX>

Query Match          18.8%; Score 185.5; DB 1; Length 252;
Best Local Similarity 36.1%; Pred. No. 2.8e-07;
Matches 44; Conservative 16; Mismatches 37; Indels 25; Gaps 2;

OY      65 GMIPGGGGNGNEPRQOPPPPEEPQAAMEGPQENMOP----- 103
        | : | : | : | : | : | : | : | : | : | : | : | : |
Db      96 GAVPPLGA--QQCSVPPIPGEGSGVLSVPQHMLPYNNVGTLSTRELQLNLNQHCR 153

OY      104 --RRRTKFTLLQVEELSEVRHHTQYPDVPTRELAENLGVEDKYRVWFKNRACRRH 161
         | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db     154 RKRRIIRTFITDEQLALENLNFQETKYPDVGTRERQLARKVHLKEEKYEWFVKRRAKWRQ 213

OY      162 QR 163
        :|
Db     214 KR 215

RESULT 3
AA2768
homeotic protein goosecoid [similarity] - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Sep-2000
C:Accession: AA2768
R:Blum, M.; Gaunt, S.J.; Cho, K.W.Y.; Steinbeisser, H.; Blumberg, B.; Bittner, D.; De Rubeis
Cell 69, 1097-1106, 1992
A>Title: Gastrulation in the mouse: the role of the homeobox gene goosecoid.
A:Reference number: A42768; MUID:92315328; PMID:1352187
A:Accession: A42768
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-256 <BLU>
A:Cross-references: GB:M85271; NID:g193895; PIDN:AAA37826.1; PID:g193896
A>Note: sequence extracted from NCBI backbone (NCBIR:108110)
C:Superfamily: homeotic protein goosecoid; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:161-217/Domain: homeobox homology <HOX>

Query Match          18.8%; Score 185.5; DB 1; Length 256;
Best Local Similarity 36.1%; Pred. No. 2.8e-07;
Matches 44; Conservative 16; Mismatches 37; Indels 25; Gaps 2;

OY      65 GMIPGGGGNGNEPRQOPPPPEEPQAAMEGPQENMOP----- 103
        | : | : | : | : | : | : | : | : | : | : | : | : |
Db     101 GAVPPLGA--QQCSVPPIPGEGSGVLSVPQHMLPYNNVGTLSTRELQLNLNQHCR 158

OY      104 --RRRTKFTLLQVEELSEVRHHTQYPDVPTRELAENLGVEDKYRVWFKNRACRRH 161
         | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db     159 RKRRIIRTFITDEQLALENLNFQETKYPDVGTRERQLARKVHLKEEKYEWFVKRRAKWRQ 218

OY      162 QR 163
        :|
Db     219 KR 220

RESULT 4
homeobox protein Pmx - mouse
I48902
N:Alternate names: homeotic protein K-2b
C:Species: Mus musculus (house mouse)
F:date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 24-Sep-1999

```

Accession: I48902; I53118; S26077
R:Kern, M.J.; Argao, E.A.; Birkmeier, E.H.; Rowe, L.B.; Potter, S.S.
Genomics 19, 334-340, 1994
A:Title: Genomic organization and chromosome localization of the murine homeobox gene
A:Reference number: I48902; MUID:94245205; PMID:7910561
A:Accession: I48902
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-217 <RES>
A:Cross-references: EMBL:U03873; NID:g460124; PIDN:AAC52139.1; PID:g460125
R:Georjani, P.; Lilly, B.; Bryson, L.J.; Wang, Y.; Sasasoo, D.A.; Olson, E.N.
Development 115, 1087-1101, 1992
A:Title: Mox1, a mesodermally restricted protein that binds an essential s
A:Reference number: I53118; MUID:93083424; PMID:1360403
A:Accession: I53118
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-217 <RES>
A:Cross-references: GB:U06502; NID:g199583; PIDN:AAA3672.1; PID:g199584
R:Kern, M.J.; Witte, D.P.; Valerius, M.T.; Aronow, B.J.; Potter, S.S.
Nucleic Acids Res. 20, 5189-5195, 1992
A:Title: A novel murine homeobox gene isolated by a tissue specific PCR cloning strat
A:Reference number: S26076; MUID:93027261; PMID:1383943
A:Accession: S26077
A:Molecule type: mRNA
A:Residues: 1-217 <RES>
A:Cross-references: EMBL:X59726
A:Experimental source: fetal heart
C:Genetics:
A:Gene: Pmx
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: alternative splicing; DNA binding; homeobox; nucleus; transcription regul
F:95-151/Domain: homeobox homology <HOX>

Query Match 18.7%; Score 184.5; DB 2; Length 217;
Best Local Similarity 29.0%; Pred. No. 2.8e-07;
Matches 49; Conservative 27; Mismatches 54; Indels 39; Gaps 5;

31 SSBGSHGCGAGPLGMLGNPNPG-----GVNHEGNKNRDGGMT-----EGG 71
Db 2 TSSYGVHLERQPALGGRISGNDTLQAKKNFSYSLDLEAGDMVAQAQDESVEGAG 61
Y 72 -----GGNGEPPRQGPPEPEPQAAMEGQPEPMQPRPRRTFTLLQVETESY 121
Db 62 RSLSEPGELTSGSDTPQOD-----NDQINSEKKKKRQRRNRTTFNSQLQALERV 112

QY 122 FRHTQYPPVPTRRLEAENIGVTEKVRVWFNKRKRRHORELILANE 170
Db 113 FERTHTPAPFVEDLARKRVNLTEARVYWFQNRRAKFRNR-AMLANK 160

RESULT 5
S26076
homeotic protein K-2a - mouse
C:Species: Mus musculus (house mouse)
C:Date: 25-Feb-1994 #sequence_revision 01-Sep-1995 #text_change 24-Sep-1999
C:Accession: S26076
R:Kern, M.J.; Witte, D.P.; Valerius, M.T.; Aronow, B.J.; Potter, S.S.
Nucleic Acids Res. 20, 5189-5195, 1992
A:Title: A novel murine homeobox gene isolated by a tissue specific PCR cloning strat
A:Reference number: S26076; MUID:93027261; PMID:1383943
A:Accession: S26076
A:Molecule type: mRNA
A:Residues: 1-245 <RES>
A:Cross-references: EMBL:X59725; NID:g51361; PIDN:CAA42410.1; PID:g51362
A:Experimental source: fetal heart
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: alternative splicing; DNA binding; homeobox; nucleus; transcription regul
F:95-151/Domain: homeobox homology <HOX>

Query Match 18.7%; Score 184.5; DB 2; Length 245;
Best Local Similarity 29.0%; Pred. No. 3.2e-07;
Matches 49; Conservative 27; Mismatches 54; Indels 39; Gaps 5;

A:Title: HES-1, a novel homeobox gene expressed by murine embryonic stem cells, identified
 A:Reference number: S35540; MUID:93087202; PMID:1360650
 A:Accession: S35540
 C:Species: preliminary; nucleic acid sequence not shown
 A:Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 116-151 <TH2>
 A:Cross-references: EMBL:L02646
 C:Genetics:
 A:Gene: HES1
 C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:109-165/Domain: homeobox homology <HOX>

Query Match 18.1%; Score 178; DB 2; Length 185;
 Best Local Similarity 33.3%; Pred. No. 7.4e-07;
 Matches 40; Conservative 17; Mismatches 35; Indels 28; Gaps 2;

OY 73 GNGE-----PQQPQPPPEEPAQAAMEGPPENNP-----R 104
 DB 50 GNSKDNPPHAPDLPSETSPCEVDHPREERAPKYENFSASSETSLKRELSWYGR 109
 OY 105 TRTRKFTLLQVEELESFVRHTQYDPVPTRELAENLGVTEDKRVWFKNRKRCRRHORE 164
 DB 110 RPRATFONOVELEENVRVNCYPCIDIREDLAQKLNIEEDRIQIWFQNRKAKKRSRRE 169

RESULT 10
 A45452
 A:Title: Transcription factor PAX3 - human (fragments)

C:Species: Homo sapiens (man)
 C:Date: 21-Sep-1993 #sequence_revision 10-May-1996 #text_change 17-Mar-1999
 C:Accession: A45452; A56744; B45452
 R:Hotch, C.F.; Milunsky, A.; Lipsky, N.; Sheffer, R.; Clarren, S.K.; Baldwin, C.T.

A:Title: Mutations in the paired domain of the human PAX3 gene cause Klein-Waardenburg
 A:Reference number: A45452; MUID:93190976; PMID:8447316
 A:Accession: A45452
 A:Molecule type: DNA
 A:Residues: 1-28, 29-306 <HOX>
 A:Note: sequence modified after extraction from NCBI backbone
 A:Note: sequence extracted from NCBI backbone (NCBIN:126845, NCBI:126846, NCBIN:126847,
 R:Macina, R.A.; Barr, F.G.; Gallili, N.; Rietman, H.C.
 A:Title: Genomic organization of the human PAX3 gene: DNA sequence analysis of the region
 A:Reference number: A56744; MUID:95301273; PMID:7782066

A:Accession: A56744
 A:Molecule type: DNA
 A:Residues: 1-28; 43-326 <MAC>
 A:Cross-references: GB:U12263
 C:Genetics:
 A:Gene: GDB:PAX3; WS1
 A:Cross-references: GDB:120495; OMIM:148820; OMIM:193500
 A:Map position: 2q35-2q35
 C:Superfamily: paired box transcription factor Pax-3; homeobox homology; paired box home
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:67-123/Domain: homeobox homology <HOX>

Query Match 18.0%; Score 177; DB 2; Length 326;
 Best Local Similarity 36.1%; Pred. No. 1.6e-06;
 Matches 48; Conservative 16; Mismatches 41; Indels 28; Gaps 5;

OY 65 GMIP---EGGGNGEPPRQPPPEEPAQAAMEG-----PQ-----PEN 100
 DB 6 GAVPRMRRPGGQVYPR--GFPLEKKAHSIDITLSERASAPOSDEGSDIDSEPDPLK 63

OY 101 MOPRTKFTLLQVEELESFVRHTQYDPVPTRELAENLGVTEDKRVWFKNRKRCRR 160
 DB 64 RKQRRSRTTFAEOLLEERAFERTHYDITRELAORAKLTARQVWFSNRRARWK 123

OY 161 H--ORELMLANEL 171
 DB 124 QAGANQLMAFNHL 136

RESULT 11
 149265
 A:Title: paired box transcription factor pax7 - mouse (fragment)

C:Species: Mus musculus (house mouse)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Jun-2000
 C:Accession: 149265
 R:Jostes, B.; Walther, C.; Gruss, P.
 Mech. Dev. 33, 27-37, 1990
 A:Title: The murine paired box gene, Pax7, is expressed specifically during the devel
 A:Reference number: 149265; MUID:91265334; PMID:1982921
 A:Accession: 149265
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-290 <RES>
 A:Cross-references: EMBL:U20792; NID:9736380; PIDN:AAA64491.1; PID:9736381
 C:Superfamily: unassigned homeobox proteins; homeobox homology; paired box homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:1-125/Domain: paired box homology (fragment) <PBH>
 F:182-238/Domain: homeobox homology <HOX>

Query Match 17.9%; Score 176.5; DB 2; Length 290;
 Best Local Similarity 33.6%; Pred. No. 1.6e-06;
 Matches 45; Conservative 13; Mismatches 43; Indels 33; Gaps 3;

OY 51 EGVNHNENG-----MNRDGMIPRGGGNGEPPRQPPPEEPAQAAMEGPOPE 99
 DB 136 EGDKKEDEKKAHSIDIGLIDCKGNRLDEGSGVSEPP-----DPL 177

OY 100 MOPRTKFTLLQVEELESFVRHTQYDPVPTRELAENLGVTEDKRVWFKNRKRCRR 159
 DB 178 RKQRRSRTTFAEOLLEERAFERTHYDITRELAORAKLTARQVWFSNRRARWK 237

OY 160 RHORELMLANELRA 173
 DB 238 KQAG---ANQLAA 247

RESULT 12
 151226

A:Title: homeodomain protein - chicken

N:Alternate names: homeotic gene paired-related protein
 C:Species: Gallus gallus (chicken)
 C:Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 17-Nov-2000
 C:Accession: 151226; I50413
 R:Kuratani, S.; Martin, J.F.; Wawersik, S.; Lilly, B.; Elchele, G.; Olson, E.N.
 Dev. Biol. 161, 357-369, 1994
 A:Title: The expression pattern of the chick homeobox gene gmxox suggests a role in p
 A:Reference number: 151226; MUID:94148118; PMID:7906232

A:Accession: 151226
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-245 <KRP>
 A:Cross-references: GB:S69088; NID:9545359; PIDN:AA29880.1; PID:9545360
 R:Nohno, T.; Koyama, E.; Miyake, F.; Taniguchi, S.; Ohuchi, H.; Saito, T.; Noji, S.
 Dev. Biol. 158, 254-264, 1993
 A:Title: The chicken homeobox gene related to Drosophila paired is predominantly expr
 A:Reference number: 150413; MUID:93321789; PMID:8101172

A:Accession: 150413
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-154; 156-245 <NOH>
 A:Cross-references: GB:D13433; NID:9222850; PIDN:BAA02695.1; PID:9222851
 C:Genetics:
 A:Gene: gmxox; Ptx-1

C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:95-151/Domain: homeobox homology <HOX>

Query Match 17.8%; Score 176; DB 2; Length 245;
 Best Local Similarity 32.9%; Pred. No. 1.4e-06;
 Matches 46; Conservative 21; Mismatches 35; Indels 38; Gaps 5;

OY 54 VNHENGMNRDGMGMI-----PEGGGNGQEPFRQOPRPEEPAQAAMEG-----POPEN 100
 Db 36 VSHLLIDLEBAGDMVAAGDEGG-----EPGRSLTSPGTTGSDTPPODN 81
 OY 101 MO-----PRTTRKFTLLQVEELESFVRHTQYPDVPTRRELAENLGATEDKVRV 150
 Db 82 DQLNSEKKKKRQRNRRTFTNSSQLALERVFERTHYPDAPFREDLARVNLTEARVQV 141
 OY 151 FKNKRRRCRRHORELMANE 170
 Db 142 FONRRAKFRNRER-AMLASK 160

RESULT 13

JC6540
 Placenta specific-homeobox protein - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 15-Oct-1999
 C:Accession: JC6540
 R:Han, Y.J.; Park, A.R.; Sung, D.Y.; Chun, J.Y.
 Gene 207, 159-166, 1998
 A:Title: Pax, a novel murine homeobox gene expressed in placenta.
 A:Reference number: JC6540; MUID:98172748; PMID:9511757
 A:Accession: JC6540
 A:Molecule type: mRNA
 A:Residues: 1-247 <HAX>
 A:Cross-references: GB:AF017453
 C:Comment: This protein is involved in controlling cell fate during embryonic development
 C:Genetics:
 A:Gene: Pax
 C:Superfamily: mouse placenta-specific homeobox protein; homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; placenta; transcription regulation
 F:151-207/Domain: homeobox homology <HOX>

Query Match 17.8%; Score 176; DB 1; Length 247;
 Best Local Similarity 34.3%; Pred. No. 1.4e-06;
 Matches 47; Conservative 23; Mismatches 41; Indels 26; Gaps 5;

OY 34 EG-HYGOGAPGLMGNNPEGVNHENGMNRDGMGMIPEGGG-NOEPROQOPRPEEPAQA 91
 Db 99 EGRHAGDGA-----SSSEDSSTLEGGQNIQOPPOEASPD----- 136
 OY 92 AMEGPOPEN-MQPTRTKFTLLQVEELESFVRHTQYPDVPTRRELAENLGATEDKVRV 149
 Db 137 SIRNHVNLRLAQLRYRTRFTHSQHLDERLFOETRYPSLARADLANMGVDECDVON 196
 OY 150 WFKNRARCRRHORELM 166
 Db 197 WFRMRALFORNRVLM 213

RESULT 14

148713
 Phox2 homeodomain protein - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 24-Sep-1999
 C:Accession: 148713
 R:Valarche, I.; Tissler-Seta, J.P.; Hirsch, M.R.; Martinez, S.; Goridis, C.; Brunet, J.F.
 Development 119, 881-896, 1993
 A:Title: The mouse homeodomain protein Phox2 regulates Ncam promoter activity in concert
 A:Reference number: 148314; MUID:94244481; PMID:7910552
 A:Accession: 148713
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-280 <RES>
 A:Cross-references: EMBL:X75014; NID:g402641; PIDN:CAA52923.1; PID:g402642
 C:Genetics:
 A:Gene: Phox2
 C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:91-147/Domain: homeobox homology <HOX>

Query Match 17.8%; Score 176; DB 2; Length 280;

Best Local Similarity 31.2%; Pred. No. 1.7e-06;
 Matches 44; Conservative 22; Mismatches 37; Indels 38; Gaps 3;

OY 23 PPOGAAASAGAHVGOGAPGLMGNNPEGVNHENGMNRDGMGMIPEGGGNOEPROQO 82
 Db 48 PCPALGSSNCALGRDHPA-----PYSAPYK-----FPPEGLHEKRQ--- 90
 OY 83 PPPEEPAQAAMGPOPENMQPTRTKFTLLQVEELESFVRHTQYPDVPTRRELAENLG 142
 Db 91 -----RRIRFTFSQQLKELERVFERTHYPDYITRELAALKIDL 129
 OY 143 TEDKVRVWFKNRARCRRHOR 163
 Db 130 TEARVQVWFONRRAKFRKOR 150

RESULT 15

JC6130
 paired box transcription factor Pax-6 - Ribbonworm
 C:Species: Lineus sanguineus (ribbonworm)
 C:Date: 16-Apr-1997 #sequence_revision 09-May-1997 #text_change 15-Oct-1999
 C:Accession: JC6130
 R:Loosli, F.; Kmita-Cunlis, M.; Gehring, W.J.
 Proc. Natl. Acad. Sci. U.S.A. 93, 2658-2663, 1996
 A:Title: Isolation of a Pax-6 homolog from the ribbonworm Lineus sanguineus.
 A:Reference number: JC6130; MUID:96181462; PMID:8610097
 A:Accession: JC6130
 A:Molecule type: DNA
 A:Residues: 1-370 <LOO>
 A:Cross-references: EMBL:X95994; NID:g1296835; PIDN:CAA64847.1; PID:e222109; PID:g129
 A:Note: The authors translated the codon GCC for residue 338 as Asp
 C:Comment: This factor is a key regulator of eye morphogenesis. It plays a role in ey
 C:Genetics:
 A:Gene: Pax-6
 A:Introns: 162/3; 236/2; 287/1; 314/3
 C:Superfamily: unassigned homeobox proteins; homeobox homology; paired box homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:47-171/Domain: paired box homology <PBH>
 F:270-326/Domain: homeobox homology <HOX>

Query Match 17.8%; Score 176; DB 2; Length 370;
 Best Local Similarity 35.1%; Pred. No. 2.3e-06;
 Matches 40; Conservative 16; Mismatches 34; Indels 24; Gaps 3;

OY 80 QPOPPPEEPAQAAMG-----POPENMQPTR-----RTKFTLLQVEEL 118
 Db 225 QPOPPISPTKKESDGHSSADSHSGDTPNGNSEEQMRIRLKRKLQNRNFTSTMAQIEAL 284
 OY 119 ESFVRHTQYPDVPTRRELAENLGATEDKVRVWFKNRARCRRHORELMANEIR 172
 Db 285 EKEFERTHYPDVFAFRERLQKIDLPARIQVWFSNRRAWRDEK---LRNQR 335

Search completed: April 28, 2003, 03:35:54
 Job time : 39 secs

This Page Blank (uspto)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2003, 19:37:10 ; Search time 23 Seconds
(Without alignments)
233.383 Million cell updates/sec

Title: US-09-867-753-2

Perfect score: 986
Sequence: 1 MARSLVHDYFVCLSYQVK.....LMLANLRADPDCCYIYVD 184

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued Patents AA:*
2: /cgn2_6/ptodata/1/laa/5A.COMB.pep:*
3: /cgn2_6/ptodata/1/laa/5B.COMB.pep:*
4: /cgn2_6/ptodata/1/laa/6A.COMB.pep:*
5: /cgn2_6/ptodata/1/laa/6B.COMB.pep:*
6: /cgn2_6/ptodata/1/laa/PCITUS.COMB.pep:*
7: /cgn2_6/ptodata/1/laa/backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	190.5	19.3	302	4	US-08-957-351-7
2	190.5	19.3	313	4	US-08-957-351-9
3	189.5	19.2	302	4	US-08-957-351-30
4	188	19.1	302	4	US-08-957-351-3
5	176	17.8	247	3	US-09-129-888-2
6	173.5	17.6	315	4	US-08-957-351-27
7	169.5	17.2	271	4	US-08-957-351-26
8	161	16.3	205	2	US-08-775-009-37
9	161	16.3	240	4	US-09-636-735A-2
10	161	16.3	240	4	US-09-636-735A-12
11	160	16.2	60	2	US-08-775-009-38
12	160	16.2	99	4	US-09-031-962D-24
13	158	16.0	436	2	US-08-958-642-4
14	158	16.0	436	3	US-08-778-394-2
15	158	16.0	436	3	US-08-778-423A-4
16	155	15.7	434	2	US-08-710-249-4
17	155	15.7	434	2	US-09-220-157A-4
18	150.5	15.3	330	2	US-08-712-948-2
19	150.5	15.3	333	2	US-08-712-948-1
20	150.5	15.3	371	2	US-08-442-809A-76
21	150	15.2	302	2	US-08-203-532F-4
22	150	15.2	302	3	US-08-950-860-16
23	150	15.2	302	4	US-09-078-165-4
24	150	15.2	302	5	PCT-US95-01882A-4
25	149	15.1	57	2	US-08-891-837B-7
26	145	14.7	303	2	US-08-203-532F-2
27	145	14.7	303	4	US-09-078-165-2

28	145	14.7	303	5	PCT-US95-01882A-2	Sequence 2, Appli
29	144.5	14.7	255	4	US-09-031-962D-4	Sequence 4, Appli
30	144	14.6	50	2	US-08-891-837B-5	Sequence 5, Appli
31	144	14.6	50	2	US-08-891-837B-6	Sequence 6, Appli
32	143.5	14.6	287	4	US-09-031-962D-2	Sequence 2, Appli
33	142.5	14.5	60	2	US-08-775-009-39	Sequence 39, Appli
34	141.5	14.4	284	2	US-08-320-148B-2	Sequence 2, Appli
35	141.5	14.4	284	3	US-08-589-028-6	Sequence 6, Appli
36	141.5	14.4	284	3	US-08-784-582-6	Sequence 6, Appli
37	141.5	14.4	284	4	US-08-785-271-6	Sequence 6, Appli
38	141.5	14.4	284	4	US-09-031-858-2	Sequence 2, Appli
39	140	14.2	147	4	US-09-605-785-336	Sequence 336, App
40	140	14.2	147	4	US-09-439-313-336	Sequence 336, App
41	140	14.2	147	4	US-09-352-616A-336	Sequence 336, App
42	140	14.2	147	4	US-09-232-149A-336	Sequence 336, App
43	138	14.0	301	4	US-09-095-117-6	Sequence 6, Appli
44	138	14.0	304	4	US-09-095-117-8	Sequence 8, Appli
45	132	13.4	66	4	US-09-031-962D-22	Sequence 22, Appli

ALIGNMENTS

```
RESULT 1
US-08-957-351-7
Sequence 7, Application US/08957351
Patent No. 6306586
GENERAL INFORMATION:
APPLICANT: Semina, Elena
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESS: FOLEY, HONG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,351
FILING DATE: 24-OCT-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UVA-024.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 302 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-957-351-7
Query Match 19.3%; Score 190.5; DB 4; Length 302;
Best Local Similarity 34.5%; Pred. No. 1.5e-11;
Matches 49; Conservative 23; Mismatches 45; Indels 25; Gaps 5;
QY 26 QLGAAASAEHGAGCAPGLGNNPREGVNHENGMNRDGGMTREGGGNOERPOPPR 85
Db 2 EFGLLSEAEAR---SPALSLIS---DACTPHP-----QLPEHGCKGQEHSD----- 40
QY 86 EEPQAAMEGRPEN-----MOPRTGRTKFTLLQVELESVFPHNQYRDPVTRRELAE 141
```

Db 41 SEKASASLPGSPEDSLKKKORRORHTFTSQQLELEATFORNRYPDMSREIAWVN 100
QY 142 VTEDKVWVFNKRRARCRHOR 163
Db 101 LTEARVWVFNKRRARCRKRER 122

RESULT 2

US-08-957-351-9
Sequence 9, Application US/08957351
Patent No. 6306586

GENERAL INFORMATION:

APPLICANT: Semina, Elena
APPLICANT: Murray, Jeffrey C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF CATARACTS
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
ZIP: 02109-2170

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,351
FILING DATE: 24-OCT-1997
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-024.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-7000
TELEFAX: 617-832-7000

INFORMATION FOR SEQ. ID NO: 9:

SEQUENCE CHARACTERISTICS:
LENGTH: 313 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-957-351-9

Query Match 19.3%; Score 190.5; DB 4; Length 313;
Best Local Similarity 34.5%; Pred. No. 1.5e-11;
Matches 49; Conservative 23; Mismatches 45; Indels 25; Gaps 5;

QY 26 QLGAASSAEHVGAGPGLGMNPNPEGVNHENGMNRDGMIEGGGNGOEPQQOPPP 85
Db 2 EFGLLSEAEAR---SPALSL---DAGTPHP-----QLPEHCKGQEHSD----- 40
QY 86 EEPQAAMEGPOPEP-----MOPRTTRTKFTLLQVEELESVFRHTQYDPVPTRELAENLG 141
Db 41 SEKASASLPGSPEDSLKKKORRORHTFTSQQLELEATFORNRYPDMSREIAWVN 100
QY 142 VTEDKVWVFNKRRARCRHOR 163
Db 101 LTEARVWVFNKRRARCRKRER 122

RESULT 3

US-08-957-351-30
Sequence 30, Application US/08957351
Patent No. 6306586

GENERAL INFORMATION:

APPLICANT: Semina, Elena
APPLICANT: Murray, Jeffrey C.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF CATARACTS
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
ZIP: 02109-2170

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,351
FILING DATE: 24-OCT-1997
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-024.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-7000
TELEFAX: 617-832-7000

INFORMATION FOR SEQ. ID NO: 30:

SEQUENCE CHARACTERISTICS:
LENGTH: 302 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-957-351-30

Query Match 19.2%; Score 189.5; DB 4; Length 302;
Best Local Similarity 34.5%; Pred. No. 1.9e-11;
Matches 49; Conservative 22; Mismatches 46; Indels 25; Gaps 5;

QY 26 QLGAASSAEHVGAGPGLGMNPNPEGVNHENGMNRDGMIEGGGNGOEPQQOPPP 85
Db 2 EFGLLSEAEAR---SPALSL---DAGTPHP-----QLPEHCKGQEHSD----- 40
QY 86 EEPQAAMEGPOPEP-----MOPRTTRTKFTLLQVEELESVFRHTQYDPVPTRELAENLG 141
Db 41 SEKASASLPGSPEDSLKKKORRORHTFTSQQLELEATFORNRYPDMSREIAWVN 100
QY 142 VTEDKVWVFNKRRARCRHOR 163
Db 101 LTEARVWVFNKRRARCRKRER 122

RESULT 4

US-08-957-351-3
Sequence 3, Application US/08957351
Patent No. 6306586

GENERAL INFORMATION:

APPLICANT: Semina, Elena
APPLICANT: Murray, Jeffrey C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF CATARACTS
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
ZIP: 02109-2170

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30


```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/957,351
: FILING DATE: 24-OCT-1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Arnold, Beth E.
: REGISTRATION NUMBER: 35,430
: REFERENCE/DOCKET NUMBER: UTA-024.01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-832-7000
: TELEFAX: 617-832-7000
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 302 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-957-351-3

```

Query Match	19.1%	Score 188	DB 4:	Length 302
Best Local Similarity	40.6%	Pred. No. 2,6e-11		
Matches	41:	Conservative	19:	Mismatches 31: Indels 10: Gaps 2
OY	67	IPGGGGNPPQPPPEPPAPAAEPPQEN----	MPPTPRKTKTLLQVEELSEVF	122
DB	28	LPHECKGQEHSD-----SEKASASLPQSGSDGSLKKQKORRKHFTSQQLDELEATF		81
OY	123	RHTQTPDYPTRELAEALGVTEDKYKVFVKFKRAKRRHR		163
DB	82	QNNRPDMSTREILAVTNLTLEAFVAVVKKRRRAKRRKEER		122

	Query Match	17.8%;	Score 176;	DB 3;	Length 247;	
	Best Local Similarity	34.3%;	Pred. No. 3.3e-10;			
	Matches 47; Conservative		23; Mismatches 41;	Indels 26;	Gaps 5;	
OY	EG-HNGOGAGPCLMGNNPEGVHENGMMRDGKMLPEGGG--NOEPROOPORPPPEPAQA	91				
Dd	EGRNAGDGA-----SSSEDSILEEGGNIDQQPPODEAASPD-----	136				
OY	AMEGFOPEN--MOPTRRTKTLLQVEELSVFRRHTOYPDVTFRRELLENIGTEDKKRV	149				
Dd	SIRNPHVNLRLAQLRFTRFTHSQLDLERLFOETRYTSLARRDLLAMWGVEDCDYON	196				
OY	WFKNKRARCRHRQRELM	166				
Dd	WFMRRALEQRNRRIYLIM	213				

```

1  APPLICANT:  Semina, Elena
2  APPLICANT:  Murray, Jeffrey C.
3  TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR THE
4  TITLE OF INVENTION:  DIAGNOSIS AND TREATMENT OF CATARACTS
5  NUMBER OF SEQUENCES:  33
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE:  FOLEY, HONG & ELIOT LLP
8  STREET:  One Post Office Square
9  CITY:  Boston
10 STATE:  MA
11 COUNTRY:  USA
12 ZIP:  02109-2170
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE:  Floppy disk
15 COMPUTER:  IBM PC compatible
16 OPERATING SYSTEM:  PC-DOS/MS-DOS
17 SOFTWARE:  PatentIn Release #1.0, Version #1.30
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER:  US/08/957,351
20 FILING DATE:  24-OCT-1997
21 CLASSIFICATION:  435
22 ATTORNEY/AGENT INFORMATION:
23 NAME:  Arnold, Beth E.
24 REGISTRATION NUMBER:  35,430
25 REFERENCE/DOCKET NUMBER:  UTA-024.01
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE:  617-832-1000
28 TELEFAX:  617-832-7000
29 INFORMATION FOR SEQ ID NO.:  27:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH:  315 amino acids
32 TYPE:  amino acid
33 STRANDEDNESS:
34 TOPOLOGY:  linear
35 MOLECULE TYPE:  protein
36
37 US-08-957-351-27

```

Query Match	17.6%	Score 173.5	DB 4	Length 315
Best Local Similarity	37.1%	Pred. No. 8.2e-10		
Matches	39	Conservative	17	Mismatches 30
				Indels 19
				Gaps 2

Qy	63	DGGMIPEGGGGQNPBRQDPPEPEPAQAAMGSPDENMKPTBRTKPFLLQVEELESVF	122
Db	69	DGGAGSACGGG-----AEDPA-----KKKKRKRRTHTFSQQLDELAEP	109
Qy	123	RHTQPDVPTRELAENLGVTEDKYRVWPKNKARCRHROREIMLT	167
		: : :	
Db	110	QNNRPDMSMREELIAVWNTLTETPRVWPKNKARVKKRTERNOOL	154
		: : :	

```

1  APPLICATION NUMBER:  US/08/957,351
2  FILING DATE:  24-OCT-1997
3  CLASSIFICATION:  435
4  ATTORNEY/AGENT INFORMATION:
5      NAME:  Arnold, Beth E.
6  REGISTRATION NUMBER:  35,430
7  REFERENCE/DOCKET NUMBER:  UTA-024.01
8  TELECOMMUNICATION INFORMATION:
9      TELEPHONE:  617-832-1000
10     TELEFAX:  617-832-7000
11 INFORMATION FOR SEQ ID NO:  26:
12     SEQUENCE CHARACTERISTICS:
13     LENGTH:  271 amino acids
14     TYPE:  amino acid
15     STRANDEDNESS:
16     TOPOLOGY:  linear
17 MOLECULE TYPE:  protein
18 US-08-957-351-26

```

```

Query Match      17.24; Score 169.5; DB 4; Length 271;
Best Local Similarity 36.2%; Pred. No. 1.7e-09;
Matches 38; Conservative 19; Mismatches 29; Indels 19; Gaps 3.

QY 59 GMRRGGMIPGGGNOBPPOPPPEPPEPQAAMEGPPENMOPRTRRKFTLLOVEL 118
   |:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 14 GLEKDKGQ-----QGKNE-----VGAEDEPSKRRQ-RRORTFTTSOOL 54

QY 119 ESVEFHTQYDPVPTRELEANGVTEDEKRVWFKKRARCRROR 163
   |:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 55 EATIORNRYPDKSTREELAVWNTLEARVWEKNNRRARWRER 99

```

RESULT 8
 US-08-775-009-37
 : Sequence 37, Application US/08775009
 Patent No. 5935783
 GENERAL INFORMATION:
 APPLICANT: Gong, Wellong
 APPLICANT: Emanuel, Beverly S.
 APPLICANT: Budarf, Marcia L.
 APPLICANT: Roe, Bruce
 TITLE OF INVENTION: NO. 5935783el Genes Mapping in the Disgeorge and
 TITLE OF INVENTION: Velocardiofacial Syndrome Minimal Critical Region
 NUMBER OF SEQUENCES: 45
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
 ADDRESSEE: No. 5935783rlis, LLP
 STREET: One Liberty Place - 46th Floor
 CITY: Philadelphia
 STATE: PA
 COUNTRY: U.S.A.
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/775,009
 FILING DATE: 27-DEC-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Trujillo, Doreen Yalko
 REGISTRATION NUMBER: 35,719
 REFERENCE/DOCKET NUMBER: CH-0681
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 568-3100
 TELEFAX: (215) 568-3439
 INFORMATION FOR SEO ID NO: 37:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 205 amino acids
 TYPE: amino acid
 TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-775-009-37

```

Query Match          16.3%; Score 161; DB 2; Length 205;
Best Local Similarity 34.8%; Pred. No. 8,6e-09;
Matches    40; Conservative   13; Mismatches   34; Indels   28; Gaps    2

OY      77 PROCGPPPEEPA-----QAAMEGQPQENMPRTR--RT 108
        ||| | | | | | | | | | | | | | | | | | | |
Db       72 PRAAPCGPPEAAGCARGALWPLRLGPAAVPLSLGADAPGSGALPGAVGCSQRTRRHHT 131

OY      109 KETTLVEELESEVFRRHQTQYPDVPTRETLAENGTEVDKTVRKFKRARCRHROR 163
         |:| :| :| :| | | | | | | | | | | | | | | |
Db       132 IFSEEDLAELELFVNQNTYPDVSTHERLAGRLRLREEREVEVFKNNRAAKMRDOKR 186

```

```

RESULT 9
US-09-636-735A-2
: Sequence 2, Application US/09636735A
: Patent No. 6416556
: GENERAL INFORMATION:
: APPLICANT: Berg, Patricia
: TITLE OF INVENTION: No. 6416956el Transcription Factor, Bp1
: FILE REFERENCE: 179.37405X00
: CURRENT APPLICATION NUMBER: US/09/636,735A
: CURRENT FILING DATE: 2000-08-11
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 2
: LENGTH: 240
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: (117)..(177)
: OTHER INFORMATION: homeobox
US-09-636-735A-2

```

```

Query Match 16.3%; Score 161; DB 4; Length 240;
Best Local Similarity 27.3%; Pred. No. 1.1e-08;
Matches 45; Conservative 30; Mismatches 52; Indels 38; Gaps 7

QY      15 SVYQYKISPT---POLGAASSAGSHV-----GGAGAG-IMGNNPE-----GGVN 55
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db       29 AAYPGLSPTTAASPNL-SYSPRYGHLISPYTEPANPEDSYLSCOQPAALSOPLCGPAE 87
QY      56 HENGNNRSGMIPESGGGGOEPRROQPOPPEPPEPAQAAMGEPQENNQPTRRRKFTLLV 115
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db       88 HPOELEAD-----SEKPRLSPEBSERRRQAPAKLRKRP-----RTIYSLDQ 129
QY      116 EELESVFRHTQYPDVPTRRLEAENLGVTEDKYKRVWFKNRRACRR 160
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db       130 OHLNRFQHTQYLAIPERAQLAAGLGTQYKWIQFNKRKRYKK 174

RESULT 10
US-09-636-735A-12
; Sequence 12, Application US/09636735A,
; Patent No. 6416956
; GENERAL INFORMATION:
; APPLICANT: Berg, Patricia
; TITLE OF INVENTION: No. 6416956el Transcription Factor, Bp1
; FILE REFERENCE: 179.37405X00
; CURRENT APPLICATION NUMBER: US/09/636, 735A
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: HELIX 1

```

LOCATION: (127)..(136)
NAME/KEY: HELIX 2 (136)
LOCATION: (144)..(154)
NAME/KEY: HELIX 3
LOCATION: (158)..(173)
US-09-636-735A-12

Query Match 16.3%; Score 161; DB 4; Length 240;
Best Local Similarity 27.3%; Pred. No. 1.1e-08;
Matches 45; Conservative 30; Mismatches 52; Indels 38; Gaps 7;

OY 15 SVYQVKISPT-----POLGAASAEHGV-----GQAPG--LNGNMP-----GGVN 55
DB 29 AATPLGSLPTTASPNL-SYSRPYGHLSTPYTPRANPGDSYLSQQPALISQPLCGPAE 87
OY 56 HENGMMNDGMIPEGGGNGNEPQQPPPEPQAAMEGPOPENMQPRTTRKFTLLQV 115
DB 88 HPELEAD-----SEKPRLSPEPSERRPQAPAKLTKRP-----RTIYSSLOL 129
OY 116 EELESVFRHTQYDPVPTRRELAEMLGVTEDKRVWFKNKRACRR 160
DB 130 QHLNQRFOHTQYALPERAQLAOLGLTQYKIMFONKRSKYK 174

RESULT 11
US-08-775-009-38
Sequence 38, Application US/08775009
Patent No. 5935783

GENERAL INFORMATION:
APPLICANT: Gong, Wellong
APPLICANT: Emanuel, Beverly S.
APPLICANT: Budarf, Marcia L.
APPLICANT: Roe, Bruce
TITLE OF INVENTION: No. 5935783el Genes Mapping in the Digeorge and
TITLE OF INVENTION: Velocardiofacial Syndrome Minimal Critical Region
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: U.S.A.
ZIP: 19103

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/775,009
FILING DATE: 27-DEC-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Trujillo, Doreen Yanko
REGISTRATION NUMBER: 35,719
REFERENCE/DOCKET NUMBER: CH-0681
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 60 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-775-009-38

Query Match 16.2%; Score 160; DB 2; Length 60;
Best Local Similarity 52.5%; Pred. No. 2.2e-09;
Matches 31; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

OY 104 RTRRTKFTLLQVEELSVFRHTQYDPVPTRRELAEMLGVTEDKRVWFKNKRACRRHQ 162
DB 2 RRRHTIFTDEQLALEMLFQETKYVDVGTREOLARKVHLREKVEVWFKNRAKWRBOK 60

RESULT 12
US-09-031-962D-24
Sequence 24, Application US/09031962D
Patent No. 6350867

GENERAL INFORMATION:
APPLICANT: Thomas C. Hart
APPLICANT: Jennifer A. Price
TITLE OF INVENTION: Methods and Compositions for Enhancing
FILE REFERENCE: WF098-18
CURRENT APPLICATION NUMBER: US/09/031,962D
CURRENT FILING DATE: 1998-02-27
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 24
LENGTH: 99
TYPE: PRT
ORGANISM: Homo sapiens
US-09-031-962D-24

Query Match 16.2%; Score 160; DB 4; Length 99;
Best Local Similarity 36.9%; Pred. No. 4.2e-09;
Matches 31; Conservative 20; Mismatches 25; Indels 8; Gaps 1;

OY 77 PROQPOPEPPEPQAAMEGPOENMQPRTTRKFTLLQVEELSVFRHTQYDPVPTREL 136
DB 1 PRLSPSERRPQAANKLRP-----RTIYSSLOLHNLNRFQHTQYALPERAQL 52
OY 137 AENLGVTEDKRVWFKNKRACRR 160
DB 53 AAGLGITQYKIMFONKRSKYK 76

RESULT 13
US-08-958-642-4
Sequence 4, Application US/08958642
Patent No. 5948623

GENERAL INFORMATION:
APPLICANT:
APPLICANT:
TITLE OF INVENTION: NOVEL METHOD FOR TESTING THE
TITLE OF INVENTION: DIFFERENTIATION STATUS IN PANCREATIC CELLS OF A MAMMAL
NUMBER OF SEQUENCES: 16
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/958,642
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/778,423
FILING DATE: December 31, 1996
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 436 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-958-642-4

Query Match 16.0%; Score 158; DB 2; Length 436;
Best Local Similarity 30.8%; Pred. No. 4.7e-08;
Matches 41; Conservative 18; Mismatches 48; Indels 26; Gaps 3;

OY 40 GAPLGMNMFEGGVNHNNGMIPGEGGNGEPRQPPPEPQAAMEGPOPE 99

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003. Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 19:24:23 ; Search time 21 Seconds

(without alignments)
363.412 Million cell updates/sec

Title: US-09-867-753-2

Perfect score: 986

Sequence: 1 MARSLVHDTVFYCLSYGVK.....IMLANELRADPDCCYIIVD 184

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	206	20.9	331	1 SHX2_HUMAN	060902 homo sapien
2	202.5	20.5	317	1 RX2_CHICK	09pvx0 gallus gall
3	202.5	20.5	331	1 SHX2_MOUSE	P70390 mus musculi
4	191.5	19.4	371	1 OTX_STRPU	026417 stronglyloe
5	190.5	19.3	302	1 PIX3_HUMAN	075364 homo sapien
6	189.5	19.2	399	1 ALX4_MOUSE	035137 mus musculi
7	188	19.1	302	1 PIX3_MOUSE	035160 mus musculi
8	187	19.0	302	1 PIX3_RAT	P81062 rattus norv
9	185.5	18.8	256	1 GSC_BOVIN	09xsk0 bos taurus
10	185.5	18.8	256	1 GSC_HUMAN	P56915 homo sapien
11	185.5	18.8	256	1 GSC_MOUSE	002591 mus musculi
12	184.5	18.7	245	1 PMX1_HUMAN	P54821 homo sapien
13	184.5	18.6	245	1 PMX1_MOUSE	P43271 mus musculi
14	183	18.6	408	1 AL_DROME	006453 drosophila
15	183	18.5	411	1 ALX4_HUMAN	09h161 homo sapien
16	182.5	18.5	364	1 OTX_PARLI	076971 paracentrot
17	182	18.5	237	1 SHX2_RAT	035750 rattus norv
18	182	18.5	240	1 DLX4_MOUSE	P70436 mus musculi
19	182	18.5	346	1 RX_HUMAN	09y2v3 homo sapien
20	181.5	18.4	325	1 OTX_MOUSE	009113 mus musculi
21	181	18.4	419	1 GSC_DROME	P53566 drosophila
22	180	18.3	299	1 CRX_HUMAN	043186 homo sapien
23	180	18.3	299	1 CRX_MOUSE	054787 mus musculi
24	179.5	18.2	317	1 PIX2_HUMAN	095697 homo sapien
25	179	18.2	342	1 RX_DROME	0911t7 rattus norv
26	178.5	18.1	873	1 RX_MOUSE	09y2q1 drosophila
27	178.5	18.1	317	1 PIX2_MOUSE	P97474 mus musculi
28	178	18.1	185	1 HEX1_MOUSE	061658 mus musculi
29	178	18.1	240	1 GSC_BRARE	P53544 brachydanio
30	178	18.0	363	1 ARX_MOUSE	035085 mus musculi
31	177	18.0	253	1 PMX2_HUMAN	095811 homo sapien
32	176.5	17.9	290	1 PAX7_MOUSE	P47239 mus musculi
33	176.5	17.9	292	1 SHOX_HUMAN	O15266 homo sapien

34	176.5	17.9	513	1 PIX1_DROME	O18400 drosophila
35	176	17.8	245	1 PMX1_CHICK	O05437 gallus gall
36	176	17.8	280	1 PMX1_MOUSE	O62066 mus musculi
37	176	17.8	281	1 PMX1_RAT	O62782 rattus norv
38	176	17.8	479	1 PAX3_HUMAN	P23760 homo sapien
39	176	17.8	479	1 PAX3_MOUSE	P24610 mus musculi
40	175	17.7	284	1 PMX1_HUMAN	O14813 homo sapien
41	174.5	17.7	245	1 GSC_CHICK	P53545 gallus gall
42	173.5	17.6	315	1 PIX1_MOUSE	P70314 mus musculi
43	173.5	17.6	333	1 PIX2_CHICK	O93385 gallus gall
44	173	17.5	247	1 PMX2_MOUSE	O06348 mus musculi
45	173	17.5	292	1 PIX3_XENLA	O918k3 xenopus lae

ALIGNMENTS

RESULT 1
SHX2_HUMAN STANDARD: PRT: 331 AA.
ID SHX2_HUMAN
AC 060902: 060903: 060465: 060467:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Short stature homeobox protein 2 (Paired-related homeobox protein SHOX) (Homeobox protein Ogl2x).
GN SHOX2 OR SHOX OR OGL2X.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Fibroblast;
RX MEDLINE=98151525; PubMed=9482898;
RA Blaschke R.J., Monaghan A.P., Schiller S., Schechinger B., Rao E., Padilla-Nash H., Ried T., Rappold G.A.;
RT "SHOX", a SHOX-related homeobox gene, is implicated in craniofacial, brain, heart, and limb development.;
RT Proc. Natl. Acad. Sci. U.S.A. 95:2406-2411(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX Strussberg R.;
RT Submitted (Mar-2001) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 116-331 FROM N.A. (ISOFORM 2).
RC TISSUE=Craniofacial;
RX MEDLINE=98133920; PubMed=9466998;
RA Semina E.V., Reiter R.S., Murray J.C.;
RT "A new human homeobox gene OGL2X is a member of the most conserved homeobox gene family and is expressed during heart development in mouse.";
RT Hum. Mol. Genet. 7:415-422(1998).
CC - FUNCTION: May be a growth regulator and have a role in specifying neural systems involved in processing somatosensory information, as well as in face and body structure formation.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - ALTERNATIVE PRODUCTS: 2 isoforms; 1/SHOX2A/SHOX (shown here) and 2/SHOX2B/SHOX2; are produced by alternative splicing.
CC - TISSUE SPECIFICITY: Expressed in heart, skeletal muscle, liver, lung, bone marrow fibroblast, pancreas and placenta.
CC - DEVELOPMENTAL STAGE: Expressed during craniofacial development as well as in heart.
CC - SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.
CC - "BICOID" SUBFAMILY.
CC - SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
CC - SIMILARITY: CONTAINS 1 OAR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL, AJ002367, CAA05341.1; ALT_INIT.
DR EMBL, AJ002368, CAA05342.1; ALT_INIT.
DR EMBL, BC008829, AAH08829.1;
DR EMBL, AF022654; AAC39662.1; ALT_INIT.
DR EMBL, AF023203; AAC39663.1;
DR HSSP; P06601; IFTL
DR TRANSFAC; T04223;
DR TRANSFAC; T04224;
DR GeneW; HGNC:10854; SHOX2.
DR MIM; 602504;
DR InterPro; IPR000047; HTH_repressr.
DR InterPro; IPR003654; Homeo_OAR.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox.1.
DR PRINTS; PR00024; HOMEBOX.
DR PRINTS; PR00031; HTHREPRESSR.
DR PRODOM; PD000010; Homeobox.1.
DR SMART; SMO0389; HOX.1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
DR PROSITE; PS50803; OAR; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Alternative splicing.
KW DNA_BIND 140 199. HOMEBOX.
FT DOMAIN 313 326 OAR.
FT DOMAIN 60 86 POLY-GLY.
FT VARSPIC 235 246 MISSING (IN ISOPOM SHOX2B).
FT CONFLICT 115 115 E -> EGKRRKTKAEVQATLLPGDAFRFL (IN REF.
FT CONFLICT 125 125 2).
FT CONFLICT 244 244 E -> D (IN REF. 1; CAA05341).
FT CONFLICT 312 312 P -> S (IN REF. 2).
FT CONFLICT 325 325 D -> N (IN REF. 2 AND 3).
FT CONFLICT 325 325 H -> L (IN REF. 3; AAC39663).
SO SEQUENCE 331 AA; 34964 MW; 55431B073B3B2250 CRC64;

Query Match 20.9%; Score 206; DB 1; Length 331;
Best Local Similarity 34.1%; Pred. No. 3e-08;
Matches 56; Conservative 22; Mismatches 50; Indels 36; Gaps 6;

QY 24 TPOLGAASSAGHVGOCAPGLMGNMNPFGVNHENGMNRDGMIPREGGGNOEP----- 77
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 53 SPVRAAGGGGGGGGGGGGGGGG-----GGYG-----GGGAGGGGSGRSPRELDM 99

QY 78 -----RQGPDP-----PE-----EPAAAMEGFQPEPNMOPRTKRTKFTLLQVEELESV 121
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 100 GAERSREPGRPLTEVSPETLKRREDKAK-GMEDEGQYTIKORSRTNFTLEQLNELRL 158

QY 122 FRHGYDPVPRRELAEMLGVTEDKVRWFKNRKARCRHREL 165
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159 FDETHYPDAFMRELSQRLGLSEARVOYVFOFNRRACKRKQENOL 202

RESULT 2
AC RX2_CHICK STANDARD; PRT; 317 AA.
OC G9PVX0;
OC 16-OCT-2001 (Rel. 40, Created)
OC 16-OCT-2001 (Rel. 40, Last sequence update)
OC 15-JUN-2002 (Rel. 41, Last annotation update)
OC Retinal homeobox protein Rx2 (crrx2).
OC RX2 OR RAX2.
OC Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
OC [1]
OC SEQUENCE FROM N.A.
OC MEDLINE=99345958; PubMed=10415362;
OC

```

```

RA Ohnchi H,Tomomari S., Itoh H., Mikawa T., Noji S. ;
RT "Identification of chick rxv/rx genes with overlapping patterns of
RL expression during early eye and brain development." ;
RL Mech. Dev. 85:193-195(1999).
CC -i- FUNCTION: Plays a critical role in eye formation by regulating the
CC initial specification of retinal cells and/or their subsequent
CC proliferation (By similarity).
CC -i- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -i- DEVELOPMENTAL STAGE: Expressed at stage 4 in the ectoderm, at
CC stage 6 in the anterior most neural plate, at stage 7 and 8 in the
CC anterior neural fold and at stage 9-10 in the evaginating optic
CC vesicles. At stage 14, highly expressed in developing retina and
CC in infundibulum region.
CC -i- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.
CC "BICOID" SUBFAMILY.
CC -i- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
CC -i- SIMILARITY: CONTAINS 1 OAR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AB020318; BAA84749.1; -
DR HSSP; P06601; 1FJL.
DR InterPro; IPR003654; Homeo_OAR.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
DR PROSITE; PS50803; OAR; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT DOMAIN 33 40 OCTAPEPTIDE MOTIF.
FT DNA_BIND 122 181 HOMEBOX.
FT DOMAIN 294 307 OAR.
FT DOMAIN 300 304 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ SEQUENCE 317 AA; 34056 MW; 1DDDF1BEC9A24910 CRC64;

Query Match 20.5%; Score 202.5; DB 1; Length 317;
Best Local Similarity 29.3%; Pred. No. 5.1e-08;
Matches 54; Conservative 28; Mismatches 61; Indels 41; Gaps 5;

QY 33 AEGHGAGCAPGLM---GNMNPGEYVNHENGNRDCGMI-----PEGGGGQGE----- 76
Db 11 AEGAFSLSPARAFSPCGNPSRLHSIEAILIGFTKKDGLCPGPPDGGCAGSAAKBAADKRGPR 70
QY 77 ---PROGPQPEPEE-----PAQAAEGPQPEENMP-----RTRRTKFT 111
Db 71 HCLPGRAPAEPPAEHQGRFQEPYCGSASPELPADGGDGDGRKSDDEQPKKKRRRRRTFTT 130
QY 112 LLQVEELSEVRRHQYPDVPTFRRELAENLGYTDEKRVVWFKKRRACRRRHORELMLANEL 171
Db 131 TYQLHELERAFEAKSHYPVYVYSREELAMKRVNLPVEVHVQVWFQVRRAKWRQREKLEVSMSKL 190
QY 172 RADP 175
Db 191 QDSP 194

RESULT 3
ID SHX2_MOUSE STANDARD; PRT; 331 AA.
AC P70390; P70369;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Short stature homeobox protein 2 (Homeobox protein Ogl2x) (OG-12)

```

DE (Paired family homeodomain protein Prx3).
GN SHOX2 OR OG12X OR PRX3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE-Embryo;
RA MEDLINE=98058757; PubMed=9371788;
RA van Schick H.S.A., Smidt M.P., Rovescalli A.C., Luijten M.,
RA van der Kleij A.A.M., Asch S., Kozak C.A., Nirenberg M.W.,
RA Burbach J.P.H.;
RT "Homeobox gene Prx3 expression in rodent brain and extraneural
RT tissues.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:12993-12998(1997).
RN [2]
RP SEQUENCE OF 116-331 AND 235-331 FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN-BALB/C; TISSUE-Liver, and Embryo;
RA MEDLINE=97008065; PubMed=8855241;
RA Rovescalli A.C., Asch S., Nirenberg M.W.;
RT "Cloning and characterization of four murine homeobox genes";
RL Proc. Natl. Acad. Sci. U.S.A. 93:10691-10696(1996).
RN [3]
RP DEVELOPMENTAL EXPRESSION.
RX MEDLINE=98133920; PubMed=9466998;
RA Semla E.V., Reiter R.S., Murray J.C.;
RT "A new human homeobox gene OG12X is a member of the most conserved
RT homeobox gene family and is expressed during heart development in
RT mouse.";
RL Hum. Mol. Genet. 7:415-422(1998).
CC -1- FUNCTION: May be a growth regulator and have a role in specifying
CC neural systems involved in processing somatosensory information,
CC as well as in face and body structure formation. May also have a
CC role in heart development.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: 1/OG12A/PRX3A (shown here) and
CC 2/OG12B/PRX3B; are produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Highly expressed in striated muscle followed
CC by liver, kidney, testis, brain, heart, lung and spleen.
CC -1- DEVELOPMENTAL STAGE: Expressed from E9 to E16 day in the heart,
CC otic region, maxillary and mandibular components of the first
CC branchial arch, nasal processes, eyelid, midbrain, medulla
CC oblongata, limbs, dorsal root ganglia and genital tubercle. Also
CC expressed in non-neuronal structures around the oral cavity and in
CC hip and shoulder regions and in mesenchyme surrounding the
CC vertebrae.
CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.
CC "BICOID" SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U66918; AAC52833.1; -;
CC EMBL: U67055; AAC52834.1; -;
CC EMBL: U65071; -; NOT_ANNOTATED_CDS.
CC EMBL: U65072; AAC52831.1; ALT_INIT.
CC EMBL: U65072; AAC52832.1; ALT_INIT.
CC HSSP: P06601; 1FJL.
CC TRANSFAC: T03310; -;
CC TRANSFAC: T04233; -;
CC MGD: MGI:1201673; Shox2.
CC InterPro: IPR003654; Homeo_OAR.
CC InterPro: IPR001356; Homeobox.
CC Pfam: PF00046; homeobox; 1.
CC PRINTS: PR00024; HOMEBOX.

DR ProDom: PD000010; Homeobox; 1.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS50071; HOMEBOX_2; 1.
DR PROSITE: PS50803; OAR; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Alternative splicing.
FT DNA_BIND 140 199 HOMEBOX.
FT DOMAIN 313 326 OAR DOMAIN.
FT DOMAIN 59 82 POLY-GLY.
FT VARSPIC 235 246 MISSING (IN ISOFORM 2).
SO SEQUENCE 331 AA; 34905 MW; D24D29E1D73A025 CRC64;
Query Match 20.5%; Score 202.5; DB 1; Length 331;
Best Local Similarity 32.5%; Pred. No. 5.3e-08;
Matches 53; Conservative 22; Mismatches 55; Indels 33; Gaps 5;
QY 24 TPOLGAASAGHVGOGAPGLMGNMNEGVMHNGNMRDGMIPGCGGNOEP----- 77
DB 52 SPAYRAAGGGGAGGAGGGGG-----GG-----GGGAGGGGAGGAGGRFVRELD 99
QY 78 -----RQOPPP-----PE-----EPAQAMGEPENMOPRRTRKFTLLQVELESVF 122
DB 100 GAERSRPEPSRFLTEVSPELDKRDKDAKMEDEGQTKIKORRSRTNFTLEQLNELRL 159
QY 123 RHQYPPVPTRELAEMLGVTEDKVRVYFKRKRRKRRHREL 165
DB 160 DETHYPPAFMRLELSORLGLSEARVQVFOFRRKCRKROENQL 202
RESULT 4
ID OTX_STRPU STANDARD; PRT; 371 AA.
AC Q26417;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUN-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein OTX (SPOTX).
GN OTX.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95180495; PubMed=7875376;
RA Gan L., Mao C.-A., Wikramanayake A., Angerer L.M., Angerer R.C.,
RA Klein W.H.;
RT "An orthodenticle-related protein from Strongylocentrotus
RT purpuratus.";
RL Dev. Biol. 167:517-528(1995).
CC -1- FUNCTION: MAY PLAY A ROLE IN ACTIVATION OF THE SPEC2A GENE, BINDS
CC TO THE TATCC MOTIF WITH HIGH SPECIFICITY. MAY HAVE ADDITIONAL
CC FUNCTIONS IN THE DEVELOPING EMBRYOS.
CC -1- SUBUNIT: BINDS DNA AS A MONOMER.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DEVELOPMENTAL STAGE: FOUND INITIALLY IN ALL CELLS OF THE CLEAVING
CC EMBRYO, BUT GRADUALLY BECOMES RESTRICTED TO ORAL ECTODERM AND
CC ENDODERM CELLS.
CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.
CC "BICOID" SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: S76899; AAB33568.1; -;
CC HSSP: P06601; 1FJL.

Query Match	Best Local Similarity	Score	DB 1	Length	DB 2	Length	DB 3	Length	DB 4	Length	DB 5	Length	DB 6	Length	DB 7	Length	DB 8	Length	DB 9	Length	DB 10	Length	DB 11	Length	DB 12	Length	DB 13	Length	DB 14	Length	DB 15	Length	DB 16	Length	DB 17	Length	DB 18	Length	DB 19	Length	DB 20	Length	DB 21	Length	DB 22	Length	DB 23	Length	DB 24	Length	DB 25	Length	DB 26	Length	DB 27	Length	DB 28	Length	DB 29	Length	DB 30	Length	DB 31	Length	DB 32	Length	DB 33	Length	DB 34	Length	DB 35	Length	DB 36	Length	DB 37	Length	DB 38	Length	DB 39	Length	DB 40	Length	DB 41	Length	DB 42	Length	DB 43	Length	DB 44	Length	DB 45	Length	DB 46	Length	DB 47	Length	DB 48	Length	DB 49	Length	DB 50	Length	DB 51	Length	DB 52	Length	DB 53	Length	DB 54	Length	DB 55	Length	DB 56	Length	DB 57	Length	DB 58	Length	DB 59	Length	DB 60	Length	DB 61	Length	DB 62	Length	DB 63	Length	DB 64	Length	DB 65	Length	DB 66	Length	DB 67	Length	DB 68	Length	DB 69	Length	DB 70	Length	DB 71	Length	DB 72	Length	DB 73	Length	DB 74	Length	DB 75	Length	DB 76	Length	DB 77	Length	DB 78	Length	DB 79	Length	DB 80	Length	DB 81	Length	DB 82	Length	DB 83	Length	DB 84	Length	DB 85	Length	DB 86	Length	DB 87	Length	DB 88	Length	DB 89	Length	DB 90	Length	DB 91	Length	DB 92	Length	DB 93	Length	DB 94	Length	DB 95	Length	DB 96	Length	DB 97	Length	DB 98	Length	DB 99	Length	DB 100	Length	DB 101	Length	DB 102	Length	DB 103	Length	DB 104	Length	DB 105	Length	DB 106	Length	DB 107	Length	DB 108	Length	DB 109	Length	DB 110	Length	DB 111	Length	DB 112	Length	DB 113	Length	DB 114	Length	DB 115	Length	DB 116	Length	DB 117	Length	DB 118	Length	DB 119	Length	DB 120	Length	DB 121	Length	DB 122	Length	DB 123	Length	DB 124	Length	DB 125	Length	DB 126	Length	DB 127	Length	DB 128	Length	DB 129	Length	DB 130	Length	DB 131	Length	DB 132	Length	DB 133	Length	DB 134	Length	DB 135	Length	DB 136	Length	DB 137	Length	DB 138	Length	DB 139	Length	DB 140	Length	DB 141	Length	DB 142	Length	DB 143	Length	DB 144	Length	DB 145	Length	DB 146	Length	DB 147	Length	DB 148	Length	DB 149	Length	DB 150	Length	DB 151	Length	DB 152	Length	DB 153	Length	DB 154	Length	DB 155	Length	DB 156	Length	DB 157	Length	DB 158	Length	DB 159	Length	DB 160	Length	DB 161	Length	DB 162	Length	DB 163	Length	DB 164	Length	DB 165	Length	DB 166	Length	DB 167	Length	DB 168	Length	DB 169	Length	DB 170	Length	DB 171	Length	DB 172	Length	DB 173	Length	DB 174	Length	DB 175	Length	DB 176	Length	DB 177	Length	DB 178	Length	DB 179	Length	DB 180	Length	DB 181	Length	DB 182	Length	DB 183	Length	DB 184	Length	DB 185	Length	DB 186	Length	DB 187	Length	DB 188	Length	DB 189	Length	DB 190	Length	DB 191	Length	DB 192	Length	DB 193	Length	DB 194	Length	DB 195	Length	DB 196	Length	DB 197	Length	DB 198	Length	DB 199	Length	DB 200	Length	DB 201	Length	DB 202	Length	DB 203	Length	DB 204	Length	DB 205	Length	DB 206	Length	DB 207	Length	DB 208	Length	DB 209	Length	DB 210	Length	DB 211	Length	DB 212	Length	DB 213	Length	DB 214	Length	DB 215	Length	DB 216	Length	DB 217	Length	DB 218	Length	DB 219	Length	DB 220	Length	DB 221	Length	DB 222	Length	DB 223	Length	DB 224	Length	DB 225	Length	DB 226	Length	DB 227	Length	DB 228	Length	DB 229	Length	DB 230	Length	DB 231	Length	DB 232
-------------	-----------------------	-------	------	--------	------	--------	------	--------	------	--------	------	--------	------	--------	------	--------	------	--------	------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

```

CC -! SIMILARITY: BELONGS TO THE PAIRED HOMEOBOX FAMILY.
CC "BICOID" SUBFAMILY
CC -! SIMILARITY: CONTAINS 1 HOMEOBOX DOMAIN.
CC -! SIMILARITY: CONTAINS 1 OAR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sb.ch).
CC -----
DR EMBL: AF041339; AAC24502.1; -.
DR EMBL: BC011642; AAH11642.1; -.
DR HSSP: P06501; IFTL.
DR TRANSFAC: T04311; -.
DR Genew: HGNC:9006; PITX3.
DR MIM: 602659; -.
DR MIM: 107250; -.
DR InterPro: IPR003654; Homeo_OAR.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox.1.
DR PRINTS: PR00024; HOMEOBOX.
DR ProDom: PD000010; Homeobox; 1.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEOBOX_1; 1.
DR PROSITE: PS50071; HOMEOBOX_2; 1.
DR PROSITE: PS50803; OAR; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Disease mutation.
FT DNA_BIND 62 121 HOMEOBOX.
FT DOMAIN 262 275 OAR.
FT DOMAIN 240 250 POLY-ALA.
FT DOMAIN 268 272 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT VARIANT 13 13 S->N (IN ADCC).
FT VAR 13 13 /FTID-VAR_003767.
FT SEQUENCE 302 AA; 31832 MW; 1E5259205ABCE87 CRC64;

Query Match 19.3%; Score 190.5; DB 1; Length 302;
Best Local Similarity 34.5%; Pred. No. 3.5e-07;
Matches 49; Conservative 23; Mismatches 45; Indels 25; Gaps 5;

QY 26 QLGAAASAEHGAGGAGPGMGCMNMBEGGVNHEGMMNRDGMGMPGEGGGNGEPRQGPQPPP 85
Db 2 EFGLLSEAEAR-----SPALSLSS-----DAGTPHP-----QLPEHGCKGQEHSD----- 40
QY 86 EEPQAAMEGQOPEN-----MQPRTKRTKFTLLQVLEELSVFPHNTQYPPVPTTRELAEINIG 141
Db 41 SEKASASLPGSGSPGEGSLKKQRQRHTFTSOLOELATFORNYPDMSTREELIAVWTN 100
QY 142 VTDEKRVVFMKRRKRCRRHOR 163
Db 101 LTEARVRYVMFKRRKRRKRER 122

RESULT 6
ALX4_MOUSE
ID ALX4_MOUSE STANDARD: PRT: 399 AA.
AC 035137;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein aristales-like 4 (ALX-4).
GN ALX4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Embryo;
CX MEDLINE=98086222; PubMed=9426253;

```


RA Qu S., Li L., Wisdom R.;
 RT "Alx-4: cDNA cloning and characterization of a novel paired-type
 RT homeodomain protein";
 RL Gene 203:217-223(1997).
 RN (2)
 RN VARIANT LST GLN-206.
 RX MEDLINE-98301426; PubMed-9636085;
 RA Qu S., Tucker S.C., Ehrlich J.S., Levorse J.M., Flaherty L.A.,
 RA Wisdom R., Vogt T.F.;
 RT "Mutations in mouse *Aristaless-1*-like4 cause Strong's luxoid
 RT polydactyly";
 RL Development 125:2711-2721(1998).
 CC -1- FUNCTION: TRANSCRIPTION FACTOR INVOLVED IN SKULL AND LIMB
 CC DEVELOPMENT.
 CC -1- SUBUNIT: BINDS DNA.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN OSTEOBLASTS, NOT EXPRESSED IN
 CC BRAIN, HEART, INTESTINE, KIDNEY, LIVER, MUSCLE, SPLEEN AND TESTIS.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED FROM E8.25 AND CONFINED TO
 CC MESENCHYMAL CELLS THROUGHOUT THE EMBRYO DEVELOPMENT. EXPRESSION IS
 CC SEEN AT SEVERAL SITES INCLUDING CRANIOFACIAL REGION, FIRST
 CC BRANCHIAL ARCH AND ANTERIOR ASPECT OF THE LIMB BUD.
 CC -1- DISEASE: DEFECTS IN ALX4 ARE THE CAUSE OF STRONG'S LUXOID (LST)
 CC PHENOTYPE. AT HETEROZYGOSITY LST IS CHARACTERIZED BY PREAXIAL
 CC ABNORMALITIES OF THE HINDFEET AND, VERY RARELY, OF THE FOREFEET.
 CC HOMOCYOTES SHOW PREAXIAL POLYDACTYLY OF ALL FOUR LIMBS,
 CC REDUCTIONS AND DUPLICATIONS OF THE RADIUS, ABSENCE OF THE TIBIA,
 CC CRANIOFACIAL DEFECTS, REDUCTION OF THE PUBIS, AND DORSAL ALOPECIA.
 CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEOBOX FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 HOMEOBOX DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF001465; AAC39943.1; -;
 DR HSSP: P06601; 1FJL.
 DR TRANSFAC: T02967; -;
 DR MGD: MGI:108359; Alx4.
 DR InterPro: IPR003654; Homeo_OAR.
 DR InterPro: IPR001356; Homeobox.
 DR Pfam: PF00046; homeobox; 1.
 DR ProDom: PD000010; Homeobox; 1.
 DR SMART: SM00389; HOX; 1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS00071; HOMEBOX_2; 1.
 DR PROSITE: PS50803; OAR; 1.
 DR Homeobox; DNA-binding; Developmental protein; Nuclear protein;
 KW Transcription regulation; Activator; Disease mutation.
 FT DNA_BIND 202 261
 FT DOMAIN 379 392
 FT VARIANT 206 206
 FT R -> Q (IN LST; ABOLISHES DNA BINDING AND
 FT TRANSCRIPTIONAL ACTIVATION).
 SQ SEQUENCE 399 AA: 42762 MW: 28DE19DDACA21D25 CRG64;
 Query Match 19.2%; Score 189.5; DB 1; Length 399;
 Best Local Similarity 30.1%; Pred. No. 5,5e-07;
 Matches 55; Conservative 22; Mismatches 49; Indels 57; Gaps 7;
 QY 23 PTPQ---LCAASSAEHVGAGGLMGNNPREGVNHENGMNDGMIPEGGGGNDPPRQ 79
 DB 95 PTPQPPAPAPPAHLYLQGA---CKTPPDPSLK-----LQSGSGHNAALQ 139
 QY 80 QP-----OPPPEPA-----OAMEGPO-----PEMNO----- 102
 DB 140 VPCYAKESNNGEPPELPDPSEPVGMDSYLSVKETGAKGQDDRAAELPSPLELTDESCK 199
 QY 103 --PRTRTKFTLLQVBELESVFRTQYPDVPTRELAENLGVTEDEVKRVWFKNKRARCR 160

DB 200 GKRRNRRTFTTSQLEBELEKVFQKTHYPDYAREQJLAMPRTDLEAVQWQFQRRKMKR 259
 QY 161 HQR 163
 DB 260 RER 262
 RESULT 7
 ID PIX3_MOUSE STANDARD; PRT; 302 AA.
 AC 035160;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUN-1998 (Rel. 36, Last annotation update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Pituitary homeobox 3 (homeobox protein PITX3).
 GN PITX3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo, and Embryonic carcinoma;
 RX MEDLINE-97472463; PubMed-9328475;
 RA Semina E.V., Reiter R.S., Murray J.C.;
 RT "Isolation of a new homeobox gene belonging to the Pitx/Rieg family:
 RT expression during lens development and mapping to the aphakia region
 RT on mouse chromosome 19.";
 RL Hum. Mol. Genet. 6:2109-2116(1997).
 RN [2]
 RP DEVELOPMENTAL EXPRESSION
 RX MEDLINE-98282096; PubMed-9620774;
 RA Semina E.V., Ferrell R.E., Mintz-Hittner H.A., Bitoun P.,
 RA Alward W.L.M., Reiter R.S., Funkhauser C., Daack-Hirsch S.,
 RA Murray J.C.;
 RT "A novel homeobox gene PITX3 is mutated in families with autosomal-
 RT dominant cataracts and ASDM.";
 RL Nat. Genet. 19:167-170(1998).
 CC -1- FUNCTION: Appears to be involved in normal eye anterior-chamber
 CC and lens development.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: Highly expressed in developing eye lens.
 CC -1- DEVELOPMENTAL STAGE: First expressed in the eye in day E10
 CC post-coitum embryos. Throughout eye development, expressed in the
 CC lens placode and forming lens pit. From day E12, also detected in
 CC the midbrain region, tongue, incisor primordia, condensing
 CC mesenchyme around the sternum and vertebrae and in the head
 CC muscles.
 CC -1- DISEASE: Mutations in PITX3 appear to be the cause of the aphakia
 CC phenotype, a recessive homozygous disease characterized by small
 CC eyes and closed eyelids.
 CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEOBOX FAMILY.
 CC "BIOID" SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 1 HOMEOBOX DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF005772; AAB87380.1; -;
 DR HSSP: P06601; 1FJL.
 DR TRANSFAC: T02666; -;
 DR MGD: MGI:1100498; Pitx3.
 DR InterPro: IPR003654; Homeo_OAR.
 DR InterPro: IPR001356; Homeobox.
 DR Pfam: PF00046; homeobox; 1.
 DR PRINTS: PR00024; HOMEBOX.

```
DR ProDom: PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
DR PROSITE; PS50803; OAR; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT DNA_BIND 62 121 HOMEBOX.
FT DOMAIN 240 250 OAR.
FT DOMAIN 268 272 POLY-ALA.
SQ SEQUENCE 302 AA; 31714 MW; EB6EF683B349264 CRC64;

Query Match
Best Local Similarity 40.6%; Pred. No. 5.3e-07;
Matches 41; Conservative 19; Mismatches 31; Indels 10; Gaps 2;

OY 67 IPEGGNGNOEPPOPEEPAQAMEGPOPEN---MOPTRRTKFTLLQVEELSEVF 122
DB 28 LPEHCCKGGEHSD-----SEKASASLPQSGSPEDSLKKQRQRTHTFSQQLDELEATF 81

OY 123 RHTQYPDVPTRELENLGVTEDEKVRVWFKNRACRRQR 163
DB 82 QNRYPDMSTREIAVWNTLTARVRFVWFKNRACRRKR 122

RESULT 8
PIX3_RAT
ID PIX3_RAT STANDARD; PRT; 302 AA.
AC P81062;
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pituitary homeobox 3 (Homeobox protein Ptx3).
GN PITX3 OR PTX3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Hypothalamus;
RX MEDLINE=96058810; PubMed=9371841;
RA Smidt M.P., van Schaik H.S.A., Lancelot C., Tremblay J.T., Cox J.J.,
  van der Kleij A.A.M., Wolterink G., Drouin J., Burbach J.P.H.;
  "A homeodomain gene Ptx3 has highly restricted brain expression in
  mesencephalic dopaminergic neurons.";
  Proc. Natl. Acad. Sci. U.S.A. 94:13305-13310(1997).
RL [2]
RN RP REVISIONS TO 66; 108; 122 AND 132-135.
RA Smidt M.P.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Appears to be involved in normal eye anterior-chamber
CC and lens development. Transcription factor which may be involved
CC in developmental determination of the mesod.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Restricted to mesencephalic dopaminergic
CC system (MESD).
CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.
CC -1- SIMILARITY: SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ011005; CAA09455.2; -
CC HSSP; P06601; 1FTL.
CC ProDom: PD000010; Homeobox; 1.
CC TRANSFAC; T04305; -
```

```
DR InterPro; IPR003654; Homeo_OAR.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom: PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
DR PROSITE; PS50803; OAR; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT DNA_BIND 62 121 HOMEBOX.
FT DOMAIN 240 250 OAR.
FT DOMAIN 268 272 POLY-ALA.
SQ SEQUENCE 302 AA; 31728 MW; DE0E81863C4AE714 CRC64;

Query Match
Best Local Similarity 40.6%; Pred. No. 5.3e-07;
Matches 41; Conservative 19; Mismatches 31; Indels 10; Gaps 2;

OY 67 IPEGGNGNOEPPOPEEPAQAMEGPOPEN---MOPTRRTKFTLLQVEELSEVF 122
DB 28 LPEHCCKGGEHSD-----SEKASASLPQSGSPEDSLKKQRQRTHTFSQQLDELEATF 81

OY 123 RHTQYPDVPTRELENLGVTEDEKVRVWFKNRACRRQR 163
DB 82 QNRYPDMSTREIAVWNTLTARVRFVWFKNRACRRKR 122

RESULT 9
CRX_BOVIN
ID CRX_BOVIN STANDARD; PRT; 299 AA.
AC Q9XSK0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cone-rod homeobox protein.
GN CRX.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Retina;
RA Zhu X., Craft C.M.;
  "Interaction of phosducin and Phlopi with CRX: potential
  transcriptional regulation function.";
  Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RL [2]
RN RP REVISIONS TO 66; 108; 122 AND 132-135.
RA Smidt M.P.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: BINDS AND TRANSCRIPTIVATES THE SEQUENCE 5'-TATC[CA]-3'
CC WHICH IS FOUND UPSTREAM OF SEVERAL PHOTORECEPTOR-SPECIFIC GENES,
CC INCLUDING THE OPSIN GENES. ESSENTIAL FOR THE MAINTENANCE OF
CC MAMMALIAN PHOTORECEPTORS.
CC -1- SUBCELLULAR LOCATION: Nuclear (by similarity).
CC -1- TISSUE SPECIFICITY: RETINA.
CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.
CC -1- SIMILARITY: SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF154123; AAD34645.1; -
CC HSSP; P06601; 1FTL.
CC TRANSFAC; T03489; -
CC InterPro; IPR001356; Homeobox; 1.
CC Pfam; PF00046; homeobox; 1.
CC ProDom: PD000010; Homeobox; 1.
CC SMART; SM00389; HOX; 1.
```

```

DR PROSITE: PS00027; HOMEBOX_1; 1.
KM Transcription regulation; Activator; Homeobox; DNA-binding;
FT DNA_BIND 39 98 HOMEBOX.
SQ SEQUENCE 299 AA; 32253 MW; 512ED6A6DAFBAC19 CRC64;

Query Match
Best Local Similarity 41.5%; Pred. No. 6.2e-07;
Matches 39; Conservative 16; Mismatches 27; Indels 12; Gaps 2;

OY 83 PPPEPAQA-AAEQPOPEMOP-----RTRRTKFTLLQVELESPFRHTQYPDV 130
DB 7 PGPHTSVNALALSGPSVDLMHRAVSPAPRKQRRRTFTTSQLEELALRAKTYQYPRV 66

OY 131 PPRRELAENLGVTEKVRVWFKRRRCRRHORE 164
DB 67 YAREVALKINLPESRVQVFKRRRAKCRQROQ 100

RESULT 10
GSC_HUMAN STANDARD; PRT; 252 AA.
ID GSC_HUMAN
AC P56915;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein goosecoid.
GN GSC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leukocyte;
RX MEDLINE=94375063; PubMed=7916327;
RA Blum M., De Robertis E.M., Kojis T., Heinzmann C., Kilsak I.,
RA Geisler D., Sparkes R.S.;
RT "Molecular cloning of the human homeobox gene goosecoid (GSC) and
RT mapping of the gene to human chromosome 14q32.1.";
RL Genomics 21:388-393(1994).
CC -1- FUNCTION: REGULATES CHORDIN (CHRD). MAY PLAY A ROLE IN SPATIAL
CC PROGRAMMING WITHIN DISCRETE EMBRYONIC FIELDS OR LINEAGE
CC COMPARTMENTS DURING ORGANOGENESIS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.
CC "BICOID" SUBFAMILY.
DR HSSP: P06601; 1FTL.
DR TRANSFAC: T04037; -.
DR Genew; HGNC:4612; GSC.
DR MIM: 138890; -.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; Homeobox.1.
DR ProDom: PD000010; Homeobox.1.
DR SMART: SM00389; HOX.1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS50071; HOMEBOX_2; 1.
KM Developmental protein; Nuclear protein; DNA-binding; Homeobox.
FT DNA_BIND 155 214 HOMEBOX.
SQ SEQUENCE 252 AA; 27853 MW; 6C0EEC48C084D323 CRC64;

Query Match
Best Local Similarity 18.8%; Score 185.5; DB 1; Length 252;
Matches 44; Conservative 16; Mismatches 37; Indels 25; Gaps 2;

OY 65 GMPGGGGNPPRQPPPEEPQAAMEGQPEMOP----- 103
DB 96 GAVPPLGA--QQSCVPTPEGEGPSVAVSPPHOMLYNVGTLSRTQLLNQLHCR 153

OY 104 -RTRRTKFTLLQVELESPFRHTQYPDVTPRRELAENLGVTEKVRVWFKRRRCRRH 161
DB 154 RRRRRRTFTDQALALENLFOETKYPRDVGTRQALARKVHLAEKVEYVFKRRRAKRRQ 213

```

```

OY 162 QR 163
DB 214 KR 215

RESULT 11
GSC_MOUSE STANDARD; PRT; 256 AA.
ID GSC_MOUSE
AC Q02591;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein goosecoid.
GN GSC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315328; PubMed=1352187;
RA Blum M., Gaunt S.J., Cho K.W.Y., Steinbeisser H., Blumberg B.,
RA Bittner D.A., de Robertis E.M.;
RT "Gastrulation in the mouse: the role of the homeobox gene goosecoid.";
RL Cell 63:1097-1106(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6;
RX MEDLINE=98079105; PubMed=94177125;
RA Danilov V., Blum M., Schweickert A., Campione M., Steinbeisser H.;
RT "Negative autoregulation of the organizer-specific homeobox gene
RT goosecoid.";
RL J. Biol. Chem. 273:627-635(1998).
CC -1- FUNCTION: GOOSECOID-EXPRESSING REGIONS OF THE GASTRULATING
CC MOUSE EGG CYLINDER HAVE ORGANIZER-LIKE ACTIVITY WHEN TRANSPLANTED
CC INTO XENOPUS EMBRYOS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: IN EARLY GASTRULATION, EXPRESSED IN THE
CC DORSAL LIP. IN LATER STAGES OF DEVELOPMENT FOUND IN HEAD, LIMBS
CC AND BODY WALL.
CC -1- INDUCTION: BY ACTIVIN.
CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.
CC "BICOID" SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M85271; AAA37826.1; -.
DR EMBL: Y13149; CAAT73611.1; -.
DR EMBL: Y13150; CAAT73612.1; -.
DR PIR: A42768; A42768.
DR HSSP: P06601; 1FTL.
DR TRANSFAC: T02047; -.
DR MGD: MGI:95841; Gsc.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; Homeobox.1.
DR ProDom: PD000010; Homeobox.1.
DR SMART: SM00389; HOX.1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS50071; HOMEBOX_2; 1.
KM Developmental protein; Nuclear protein; DNA-binding; Homeobox.
FT DNA_BIND 160 219 HOMEBOX.
SQ SEQUENCE 256 AA; 27979 MW; 3639FB059AC3DB9E CRC64;

Query Match
Best Local Similarity 18.8%; Score 185.5; DB 1; Length 256;
Matches 44; Conservative 16; Mismatches 37; Indels 25; Gaps 2;

```


CC -1- FUNCTION: ACTS AS A TRANSCRIPTIONAL REGULATOR OF MUSCLE CREATINE
 CC KINASE (MCK) AND SO HAS A ROLE IN THE ESTABLISHMENT OF DIVERSE
 CC MESODERMAL MUSCLE TYPES. THE PROTEIN BINDS TO AN A/T-RICH ELEMENT
 CC IN THE MUSCLE CREATINE ENHANCER.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; PMX1-A AND PMX1-B (SHOWN
 CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN MESODERMALLY DERIVED
 CC CELL TYPES. DURING EMBRYOGENESIS, HIGHEST LEVELS OF EXPRESSION
 CC ARE FOUND IN THE MESENCHYME AND PRECARDIAC ELEMENTS OF THE FACE
 CC AND HIND LIMBS. IN THE ADULT, EXPRESSION IS RESTRICTED TO SKELETAL
 CC MUSCLE, HEART AND UTERUS.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING CARDIOGENESIS.
 CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEOBOX FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: L06502; AAC39672.1; -
 CC EMBL: U03873; AAC52139.1; -
 CC EMBL: X59725; CAA42410.1; -
 CC EMBL: S82911; AAB46839.1; -
 CC HSSP: P06601; 1FTJL.
 CC TRANSFAC: T02061; -
 CC TRANSFAC: T02061; -
 CC TRANSFAC: T02966; -
 CC MGD: MGI:97712; Ptx1.
 CC InterPro: IPR003654; Homeo_OAR.
 CC InterPro: IPR001356; Homeobox.
 CC Pfam: PF00046; homeobox; 1.
 CC ProDom: PD000010; Homeobox; 1.
 CC SMART: SM00389; HOX; 1.
 CC PROSITE: PS00027; HOMEBOX_1; 1.
 CC PROSITE: PS50071; HOMEBOX_2; 1.
 CC PROSITE: PS50803; OAR; 1.
 CC KMW Homeobox: DNA-binding; Developmental protein; Nuclear protein;
 CC Alternative splicing; Phosphorylation.
 CC FT DNA_BIND 94 153 HOMEBOX.
 CC FT DOMAIN 222 235 OAR.
 CC FT MOD_RES 197 197 PHOSPHORYLATION (POTENTIAL).
 CC FT VARSPPLIC 200 245 SAMATYSATCANNSPAQGINMANSTIRLAKESLORNO
 CC VPTVN -> RSSSLPRCCHGHLNGF (IN ISOFORM
 CC PMX1-A).
 CC SO SEQUENCE 245 AA; 27269 MW; FFE67FDD856E1115 CRC64;
 CC
 CC Query Match. 18.7%; Score 184.5; DB 1; Length 245;
 CC Best Local Similarity 29.0%; Pred. No. 7.7e-07;
 CC Matches 49; Conservative 27; Mismatches 54; Indels 39; Gaps 5;
 CC
 CC QY 31 SSAEGHVGCGAPLGMNMPG-----GVNHEMNANDGKIP-----EGG 71
 CC DB 2 TSYGVHLEKQPLGRLDSPNLDTLQAKKNVSVHLDLLEBAGVVAQAODESVEAG 61
 CC QY 72 -----GNGQEPROOPPPPEPPAQAAMEGPENNNPRTPTKFTLLVEELESV 121
 CC DB 62 RSLLESPLGLTSGSDTPQD-----NDQLNSEKKKKKKRNRKNTTFSSQLQALERV 112
 CC QY 122 FRITQPDVPTRRLEAENIGVEDKYRVWFKNRARRRHRRLMLANE 170
 CC DB 113 FERTHPDAFVREDLARVNLTEARVQVFNORAKFRNRER-AMLANK 160
 CC
 CC RESULT 14
 CC AL_DROME
 CC ID AL_DROME
 CC AC 006453;
 CC
 CC STANDARD: PRT; 408 AA.
 CC

DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Homeobox protein aristaless.
 GN AL.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Braachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=93138380; PubMed=8093690;
 RA Schneitz K., Spielmann P., Noll M.;
 RT "Molecular genetics of aristaless, a prd-type homeo box gene involved
 RT in the morphogenesis of proximal and distal pattern elements in a
 RT subset of appendages in Drosophila.";
 RL Genes Dev. 7:114-129(1993).
 CC -1- FUNCTION: INVOLVED IN THE MORPHOGENESIS OF PROXIMAL AND DISTAL
 CC PATTERN ELEMENTS IN A SUBSET OF APPENDAGES. APPEARS ALSO TO HAVE A
 CC ROLE IN EARLY IMAGINAL DISK DEVELOPMENT. ARISTALESS MUTANTS
 CC DISPLAY A REDUCTION IN SIZE OF THE ARISTA AND SCUTELLUM, A
 CC REDUCTION OR COMPLETE ABSENCE OF THE TARSAL CLAWS, IRREGULARITIES
 CC OF THE STERNOPLURAL BRISTLES AND OF THE WING VEIN, AND A BENDING
 CC OF THE WING BLADE.
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- TISSUE SPECIFICITY: EXPRESSED DURING EMBRYONIC DEVELOPMENT, IN
 CC DISTINCT EPIDERMAL PATTERNS IN HEAD, THORAX AND ABOOMEN AND IN AN
 CC ENDODERMAL PATTERN IN THE ANTERIOR INTESTINAL TRACT.
 CC -1- DEVELOPMENTAL STAGE: FIRST EXPRESSED IN 4-8 HR-OLD EMBRYOS, PEAKS
 CC IN 8-12 HR-OLD EMBRYOS AND CONTINUES THROUGH DEVELOPMENT UP TO
 CC LATE LARVAL STAGE. IN THE HEAD REGION, DETECTED AT STAGE 10 IN THE
 CC MAXILLARY AND LABIAL SEGMENT PRIMORDIA, AND IN LATER STAGES, IN
 CC THE PROSPECTIVE ANTENNAL AND MANDIBULAR SEGMENT. IN THE EPIDERMIS,
 CC EXPRESSION IS SEEN FROM STAGE 11 IN THORACIC AND ABDOMINAL LATERAL
 CC PATCHES. EXPRESSION IN THE INTESTINAL TRACT BEGINS AT STAGE 13,
 CC CONTINUES THROUGH STAGES 14 AND 15 IN THE ENDODERM OF THE ANTERIOR
 CC MIDGUT AND AT STAGE 16, IS FOUND IN THE POSTERIOR END. EXPRESSION
 CC IN THE IMAGINAL DISKS IS SEEN FROM LATE THIRD-INSTAR LARVAE IN
 CC THE PROSPECTIVE THORAX, CLAW ORGAN, ANTENNA, SCUTELLUM AND WING
 CC BLADE.
 CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEOBOX FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 HOMEOBOX DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: L08401; AAA28840.1; -
 CC HSSP: P06601; 1FTJL.
 CC FlyBase: FBgn0000061; al.
 CC InterPro: IPR000047; HTH_Repressor.
 CC InterPro: IPR003654; Homeo_OAR.
 CC InterPro: IPR001356; Homeobox.
 CC Pfam: PF00046; homeobox; 1.
 CC PRINTS: PR00024; HOMEBOX.
 CC PRINTS: PR00031; HTHREPRESSR.
 CC ProDom: PD000010; Homeobox; 1.
 CC SMART: SM00389; HOX; 1.
 CC PROSITE: PS00027; HOMEBOX_1; 1.
 CC PROSITE: PS50071; HOMEBOX_2; 1.
 CC PROSITE: PS50803; OAR; 1.
 CC KMW Homeobox: DNA-binding; Developmental protein; Nuclear protein.
 CC FT DNA_BIND 85 144 HOMEBOX.
 CC FT DOMAIN 378 391 OAR.
 CC FT DOMAIN 251 283 PRO-RICH.
 CC FT DOMAIN 290 360 GLN/PRO-RICH.
 CC

Mon Sep 8 15:07:38 2003

us-09-867-753-2.rsp

Page 11

```
QY      143 TEDKVRVWEKNNKRARCRRHQ 163
          || :|||:|:|: | :|
Db      254 TEARVQVWFQNNRRRAKWKRRER 274
```

Search completed: April 28, 2003, 21:14:05
Job time : 24 secs

This Page Blank (uspto)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 28, 2003, 16:24:32 ; Search time 54 Seconds

(without alignments)
3276.898 Million cell updates/sec

Title: US-09-867-753-1

Perfect score: 577

Sequence: 1 tcaacatcagcgctccag.....tctacatcgtcgtgactag 577

Scoring table: IDENTITY_NUC

Searched: Gapop 10.0 , Gapext 1.0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA :
1: /cgn2_6/prodata1/1na/5A_COMB.seq : *
2: /cgn2_6/prodata1/1na/5B_COMB.seq : *
3: /cgn2_6/prodata1/1na/5A_COMB.seq : *
4: /cgn2_6/prodata1/1na/5B_COMB.seq : *
5: /cgn2_6/prodata1/1na/PCRTUS_COMB.seq : *
6: /cgn2_6/prodata1/1na/backfile1.seq : *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	71	12.3	877	US-09-129-888-1	Sequence 1, Appl
2	56.8	9.8	4080	US-08-710-249-3	Sequence 3, Appl
3	56.8	9.8	4080	US-09-220-157A-3	Sequence 3, Appl
4	56.4	9.8	906	US-08-957-351-6	Sequence 6, Appl
5	56.4	9.8	1017	US-08-957-351-5	Sequence 5, Appl
6	55.4	9.6	2481	US-08-958-642-3	Sequence 3, Appl
7	55.4	9.6	2481	US-08-778-394-1	Sequence 1, Appl
8	55.4	9.6	2481	US-08-778-423A-3	Sequence 3, Appl
9	54.6	9.5	906	US-08-957-351-2	Sequence 2, Appl
10	54.6	9.5	1392	US-08-957-351-1	Sequence 1, Appl
11	52	9.0	1251	US-09-636-735A-1	Sequence 1, Appl
12	51.6	8.9	1275	US-08-958-642-1	Sequence 1, Appl
13	51.6	8.9	1275	US-08-958-642-3	Sequence 1, Appl
14	51.6	8.9	1275	US-08-778-394-3	Sequence 1, Appl
15	48.2	8.4	838	US-07-590-894C-1	Sequence 1, Appl
16	48.2	8.4	7218	US-08-332-463-14	Sequence 14, Appl
17	47.8	8.3	410	US-09-319-648-45	Sequence 45, Appl
18	43.8	7.6	157	US-09-277-078-41	Sequence 41, Appl
19	43.4	7.5	2338	US-08-425-069-1	Sequence 1, Appl
20	43.4	7.5	2338	US-08-317-844B-1	Sequence 1, Appl
21	42.8	7.4	786	US-08-403-379A-2	Sequence 2, Appl
22	42.8	7.4	786	US-08-929-414-2	Sequence 2, Appl
23	42.4	7.3	1223	US-08-957-351-4	Sequence 4, Appl
24	42.4	7.3	1223	US-08-957-351-29	Sequence 29, Appl
25	42.4	7.3	1240	US-08-957-351-8	Sequence 8, Appl
26	41.8	7.2	1500	US-09-593-711A-10	Sequence 10, Appl
27	41.4	7.2	245	US-09-117-121-27	Sequence 27, Appl

28	41.2	7.1	467	US-08-712-948-6	Sequence 6, Appl
29	41	7.1	1021	US-09-095-117-5	Sequence 5, Appl
30	41	7.1	1030	US-09-095-117-7	Sequence 7, Appl
31	41	7.1	1995	US-08-425-069-3	Sequence 3, Appl
32	41	7.1	1995	US-08-317-844B-3	Sequence 3, Appl
33	41	7.1	4403765	US-09-103-840A-2	Sequence 2, Appl
34	40.2	7.0	985	5215895-1	Patent No. 5215895
35	39.8	6.9	432	US-09-397-787-191	Sequence 191, App
36	39.8	6.9	15378	US-08-785-420-1	Sequence 1, Appl
37	39.4	6.8	409	US-09-319-648-22	Sequence 22, Appl
38	39.4	6.8	432	US-08-642-255-48	Sequence 48, Appl
39	39.4	6.8	756	US-08-642-255-50	Sequence 50, Appl
40	39.4	6.8	4403765	US-09-103-840A-2	Sequence 2, Appl
41	39.2	6.8	477	US-09-135-994-1	Sequence 1, Appl
42	39.2	6.8	4425	US-08-749-169A-1	Sequence 1, Appl
43	39.2	6.8	4425	US-09-130-032A-1	Sequence 1, Appl
44	38.8	6.7	292	US-09-117-121-29	Sequence 29, Appl
45	38.8	6.7	292	US-09-344-529-8	Sequence 8, Appl

ALIGNMENTS

```
RESULT 1
US-09-129-888-1
: Sequence 1, Application US/09129888B
: Patent No. 6063912
: GENERAL INFORMATION:
: APPLICANT: CHUN, Jong Yoon
: TITLE OF INVENTION: Placenta trophoblast-specific gene
: FILE REFERENCE: 1942/29
: CURRENT APPLICATION NUMBER: US/09/129, 888B
: CURRENT FILING DATE: 1998-08-06
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: Wordperfect 6.1 Windows
: SEQ ID NO 1
: LENGTH: 877
: TYPE: DNA
: ORGANISM: mouse
US-09-129-888-1

Query Match      12.3%   Score 71;   DB 3;   Length 877;
Best Local Similarity 56.7%   Pred. No. 4.6e+09;
Matches 131;   Conservative 0;   Mismatches 100;   Indels 0;   Gaps 0;

QY 326 CAGCCAGCAACTCGCCGACGAGTTCACGCTGTCGAGTGAGAGCTGGAAGTGT 385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 488 CAACTGCGGTACAGACGACGACGACGACGACGACGACGACGACGACGACG 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 386 TTCGACACACTCAATACCTGATGTCGCCACAAGAGGAATTCGCCGAATTA 445
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 548 TTCGAAGAGACTGCTACCCAGCTTCGACGACGAGAGGATCTTCACATGAT 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 446 GTGACTGAAGCAAGTCGCGGTTGTTGAAGATTAAGGCCACATGATGACG 505
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 608 GTGATGAATGTGATGTCGACGAATTTGTTGATGATGAGAGACCTTTTCCA 667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 506 CAGGAGAAATTAATGTCGCCAATGATGATGATGATGATGATGATGATGAT 556
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 668 AGGAGATGCTGATGTTCTGTGGAAGTCCGCTCTTCCCAAGAGACGACTCT 718
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 2
US-08-710-249-3
: Sequence 3, Application US/08710249
: Patent No. 5858777
: GENERAL INFORMATION:
: APPLICANT: Villeponteau, Bryant
: APPLICANT: Feng, Junli
: APPLICANT: Andrews, William H.
: APPLICANT: Adams, Robert R.
: TITLE OF INVENTION: Methods and Reagents for Regulating
```



```

? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/958,642
? FILING DATE:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/778,423
? FILING DATE: December 31, 1996
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2481 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: other nucleic acid
? DESCRIPTION: /desc = "Oligonucleotide"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 163..1470
? US-08-958-642-3

```

Query Match	9.68;	Score 55.4;	DB 2;	Length 2481;
Best Local Similarity	55.48;	Pred. NO. 4.7e-05;		
Matches 107; Conservative	0;	Mismatches 86;	Indels 0;	Gaps 0;

QY	319	GACATATCCACACACAGAACTCGGGCAGCAAGATTCACGCTGTGGTGGAGAGACGTGA	378
Db	822	GAACGGAGACGTCCAAAGAAATAGAACATCTTTTACCCAGACACAGATTGAGGCTCTGGA	881
QY	379	AAGTGTTCGCGACACATCAATCCCTGATGTGCCACAGAAAGGAACCTTGCCGAAA	438
Db	882	GAAGAGCTTTGAGAGAGACCCATTTATCCAGATGTGTTTGCCCGGGAAGAACTAGCAGCCA	941
QY	439	CTTAGGTGTGACATGAGACCAAAAGTCGGGGTTTGTTTAAAGATTAAGAGGCCAGATGTAG	498
Db	942	AATGATCTACTCCGAGACAGAAATACAGATGTGTTTCTTAATCCAGAGGCCAAATGGAG	1001
QY	499	GGCAGATCAGAGA	511
Db	1002	AAGAGAGAGAAA	1014

RESULT 7
 US-08-778-394-1
 ; Sequence 1, Application US/08778394
 ; Patent No. 6028184
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; APPLICANT:
 ; TITLE OF INVENTION: NOVEL METHOD FOR TESTING THE
 ; TITLE OF INVENTION: DIFFERENTIATION STATUS IN PANCREATIC CELLS OF A MAMMAL
 ; NUMBER OF SEQUENCES: 4

```

1 COMPILED READABLE: F04A.
2 MEDION TYPE: Floppy disk
3 COMPUTER: IBM PC compatible
4 OPERATING SYSTEM: PC-DOS/MS-DOS
5 SOFTWARE: PatentIn Release #1.0, Vers
6 CURRENT APPLICATION DATA:
7 APPLICATION NUMBER: US/08/778,394
8 FILING DATE:
9 CLASSIFICATION: 514
10 INFORMATION FOR SEQ ID NO: 1:
11 SEQUENCE CHARACTERISTICS:
12 LENGTH: 2481 base pairs
13 TYPE: nucleic acid
14 STRANDEDNESS: single
15 TOPOLOGY: linear
16 MOLECULE TYPE: other nucleic acid
17 DESCRIPTION: /desc = "Oligonucleotide
18 FEATURE:
19 NAME/KEY: CDS
20 LOCATION: 163..1470
21 OS-08-778-394-1

```

Query Match	9.6%	Score 55.4;	DB 3;	Length 2481;
Best Local Similarity	55.4%;	Pred. NO. 4.7e-05;		
Matches 107;	Conservative 0;	Mismatches 86;	Indels 0;	Gaps 0;

OY	319	GAACTATCGACAGCAAGCACTGGGCGAAGAAGTTACACGTCTGTTGCAGGTGAGAGACTCTGA	378
Db	822	GAAACCCGAAGCTGCAGAAATATGAACATCTTTTACCACAAGCAGCATTTGAGGCTCTGGA	881
OY	379	AAGTGTTTTCCGACACACATCAATACCTCTGATGTGCCCAAGAAGGGAATTGCCGAAA	438
Db	882	GAAAGAGTTTGAAGAGGAGCCCATTAATCCAGATGTGTTTGCCCGGGAAGAAGCTAGCAGCCAA	941
OY	439	CTTAGTGTCACATCGAAGCAAAGTCGGGGTTTGCTTTAAAGAAATAAAGGGCCAGATGTAG	498
Db	942	AATGATCACTCACTGAMACCAAGAAATACAGATGATGTTTCTAATCGAAGGGCCCAATATGAG	1001
OY	499	GCGCATCAAGAGA	511
Db	1002	AAGAGAAAGAGAA	1014

```

      RESULT 8
      US-08-778-423A-3
      Sequence 3, Application US/08778423A
      Patent No. 6071697
      GENERAL INFORMATION:
      APPLICANT:
      APPLICANT:
      TITLE OF INVENTION: NOVEL METHOD FOR TESTING THE
      TITLE OF INVENTION: DIFFERENTIATION STATUS IN PANCREATIC CELLS OF A MAMMAL
      NUMBER OF SEQUENCES: 16
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/778,423A
      FILING DATE: December 31, 1996
      INFORMATION FOR SEQ ID NO: 3:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 2481 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: other nucleic acid
      DESCRIPTION: /desc = "oligonucleotide"
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 163..1470
      US-08-778-423A-3
  
```

Query Match	9.6%	Score 55.4	DB 3	Length 2481
Best Local Similarity	55.4%	Pred. No. 4.7e-05		
Matches 107	Conservative 0	Mismatches 86	Indels 0	Gaps 0
QY 319	GAACATGCGACGACCAACTCGCGCCACGAAAGTTCACGGTGTGACAGGTGGAGAGCTGGA	378		
Db 822	GAAGCGGAAGCTGCCAAAGAAATAGAACATCTTTTACCAGAGCAGATTTGAGGCTCTGGA	881		
QY 379	AAGTGTTCCTCCGACACTCAATACCCTTGATGTGCGCCACAAGAGGGAATCTTGCCGAAAA	438		
Db 882	GAAGAGATTTTGAAGAGACCATTATCCAGATGTGTTTCCCGGGAAAGACTGACGCCAA	941		
QY 439	CTTAGGTGTGACTGAGAACAAAGTCGCGGTTTGTTTAAAGATTAAGGGCCAGATGTAG	498		
Db 942	AATAGATCTTACTCTGAGAGCAAGATACAGTATGTGTTTCTATATCAAGGGCCAAATGGAG	1001		
QY 499	GCGACATCAGAGA	511		
Db 1002	AAGAGAAAGAGAAA	1014		


```

: TELEPHONE: 713-651-5587
: TELEFAX: 713-651-5246
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 838 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: ORIGINAL SOURCE:
: STRAIN: AKR1 Jackson
: INDIVIDUAL ISOLATE: SL12 cell line
: DEVELOPMENTAL STAGE: Bone marrow-adult
: TISSUE TYPE: Lymphoma
: CELL TYPE: T-cell
: CELL LINE: SL12.4 clone
: FEATURE:
: OTHER INFORMATION: N at nucleotide 838 represents a polyA string
US-07-550-894C-1

Query Match      8.4%; Score 48.2; DB 1; Length 838;
Best Local Similarity 49.8%; Pred. No. 0.0023;
Matches 150; Conservative 0; Mismatches 148; Indels 3; Gaps 1;

QY 256 GCAGCAGCCGACGCCCCCGGAGAGCCGCGCCAGCGCGCCATGAGAGGTCCCGACGC 315
   ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 366 GGAGCGAGGACAAATGAGCCAGTTGCTGAGGGGCACCTGAGAGCCAGGAAATGAAATCC 425

QY 316 CGAGAACATGACGAGCCGAACTCGGCGGACGAGAAATGCAACGCTTGTCAGGTGAGAGAGCT 375
   || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 426 TGGGGGTAGGAGATGCCCTCCAGGCGCTGAGGTGCGCCGCAATCGACTGAGGAACT 485

QY 376 GGAAGTCTTTCCGACACACTCAATACCTGATGTGCCACAAAGAGGAACTTGCCGA 435
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 486 GGAGTCAATTTGCAGCGCACAATTCCTTGATGT---CCCAAGGAGGATCTTGATAG 542

QY 436 AAATCTAGGTGACTGAAGACAAAGTCGGGTTTGTTTAAGATTAAGAGGCCAGATG 495
   || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 543 ACTGATGATGCTCTGTGTCTGTCAGAGTGCAGAAATGGTTTAAGATCAGAGGGCGTGC 602

QY 496 TAGGCGATCATGAGAGAAATTAATGCTGCGCAATGAACTAGCTGCTGAACCCAGAGAGCTG 555
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 603 CAGAGAAACAGAGAGGAGGCGACACACAGTCCCTGAACATTTTAGAGGAACATTCGAGTG 662

QY 556 T 556
   |
Db 663 T 663

Search completed: April 28, 2003, 19:25:28
Job time : 72 secs

```

This Page Blank (uspto)